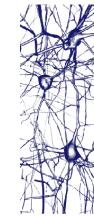




ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE



Leveraging heterogeneous systems and deep memory hierarchies for brain tissue modeling

Tier 1 ALCF Theta Early Science Program

Fabien Delalondre

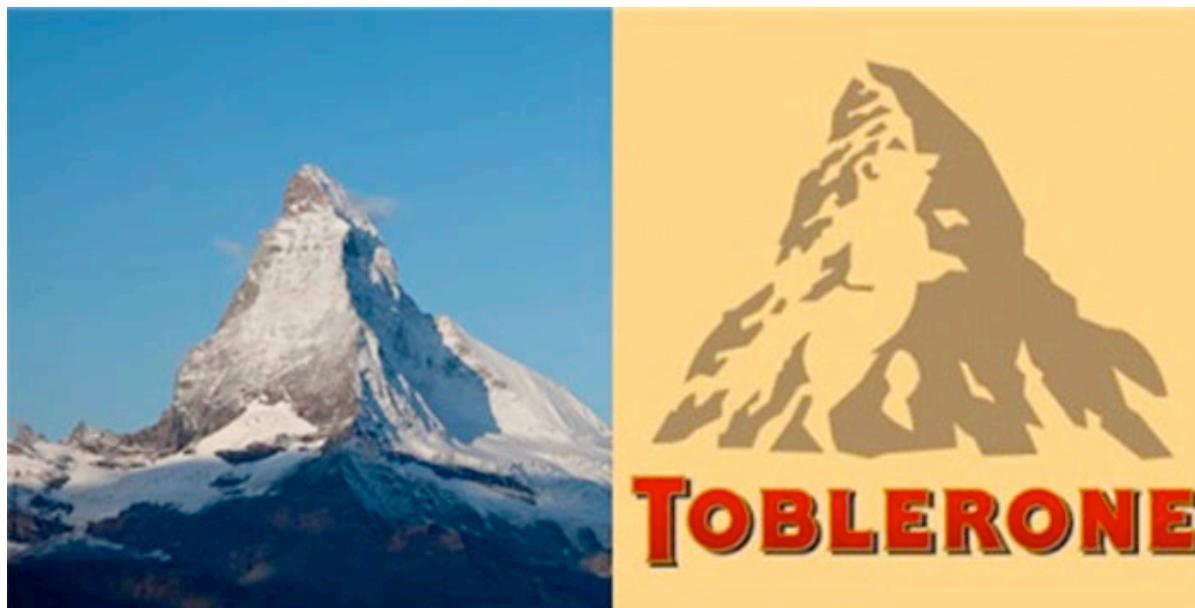
Blue Brain Project, HPC Team Manager

- **Blue Brain Project (BBP) & Human Brain Project (BBP) introduction**
- On node portable performance optimization
- Future R&D Directions



- ~90 people & 25 nationalities, expecting 110-120 people in next 2-4 years (~20M budget/year)
- Operates heterogeneous infrastructure (4 rack Blue Gene/Q, clusters, storage, private cloud, volunteer computing, desktops, ...)
- More than 50 applications written in python, java, C/C++, ... organized in complex workflows
- Extensive software engineering practices & infrastructures (more than just continuous integration ...)

Toblerone Chocolate...



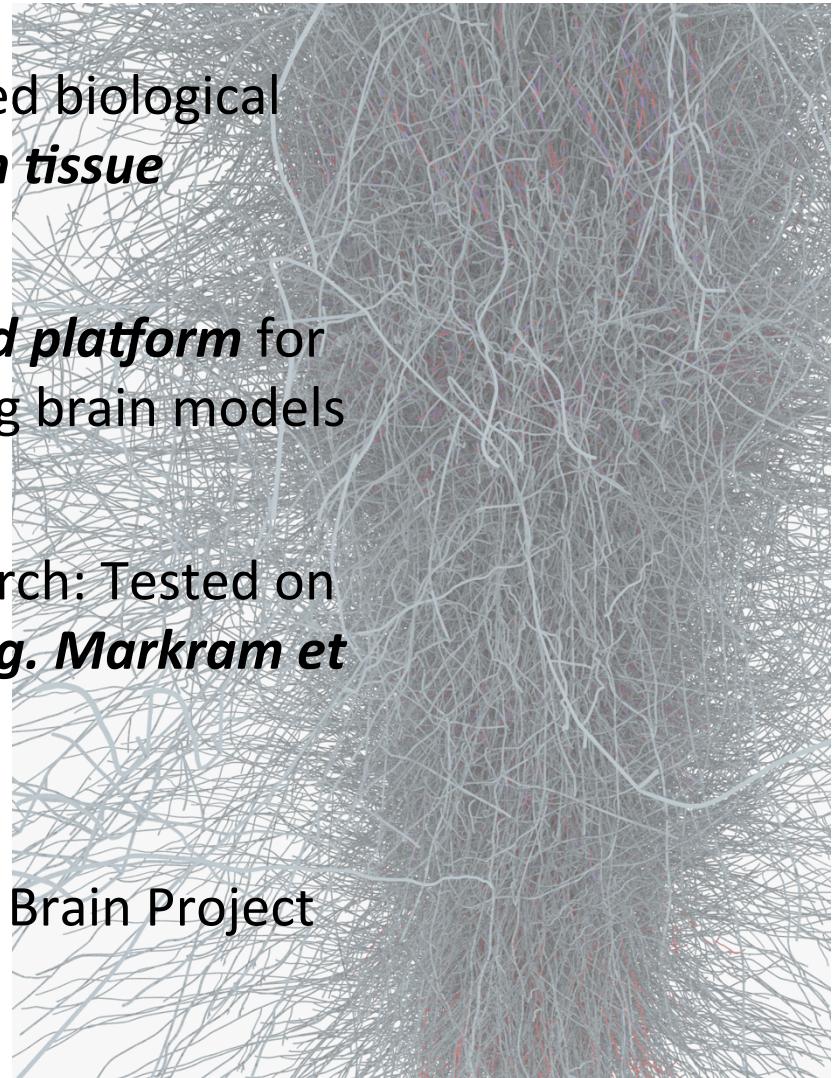
Matterhorn



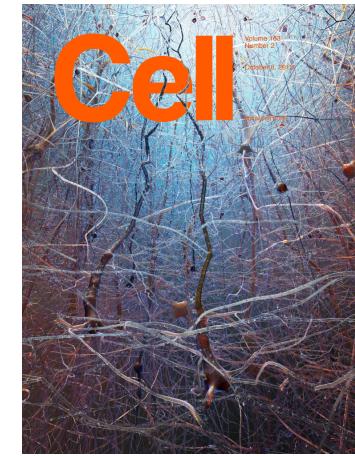
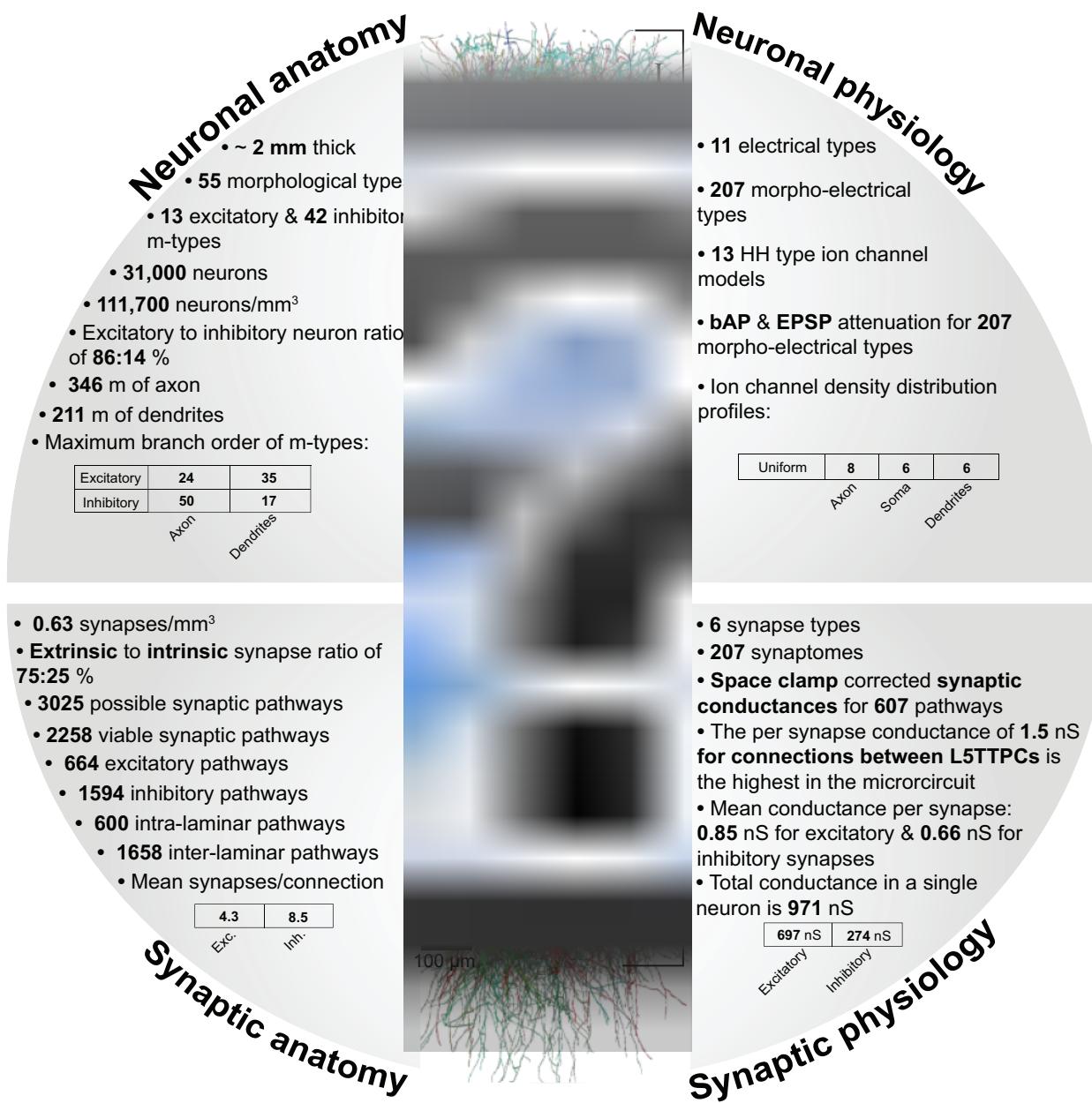
Help Yourself !

Blue Brain Project Development Strategy

- Pioneers strategy to integrate fragmented biological knowledge into ***unifying models of brain tissue***
- Develops ***unique (super)computer-based platform*** for building, simulating & evaluating unifying brain models
- Use platform for simulation-based research: Tested on a ***rat's neocortical microcircuit*** → see e.g. ***Markram et al. 2015 (Cell)***
- Simulation core of the European Human Brain Project



Data-Driven Modeling & Simulation



Markram et al, Cell 2015

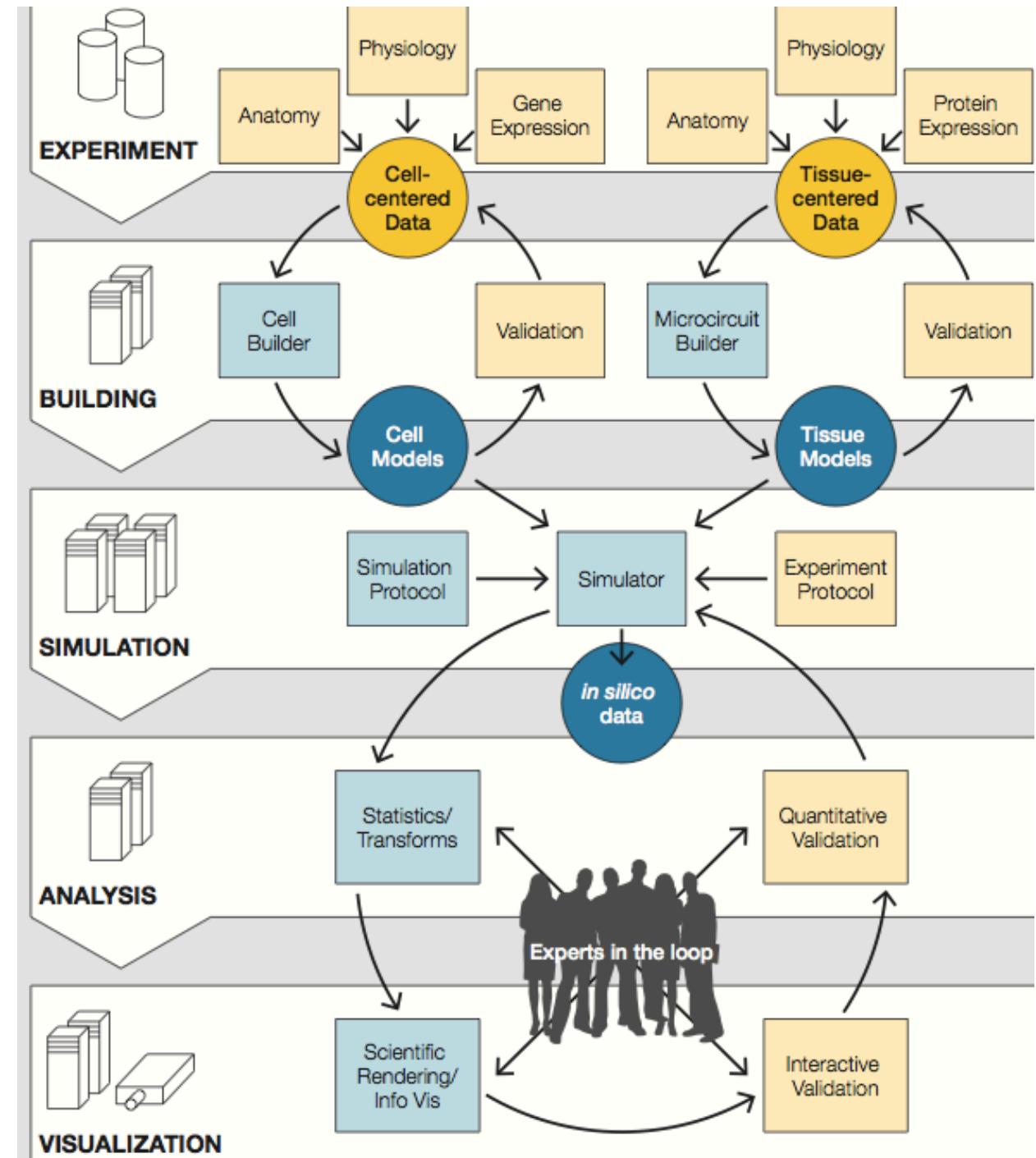
<https://bbp.epfl.ch/nmc-portal>

- 0.28mm²
- 31'000 neurons
- 207 morpho-electrical types
- 31'628 types of connections
- 40 million intrinsic synapses
- 141 million extrinsic synapses

Co-designed Heterogeneous ICT Facility

Relevant publications:

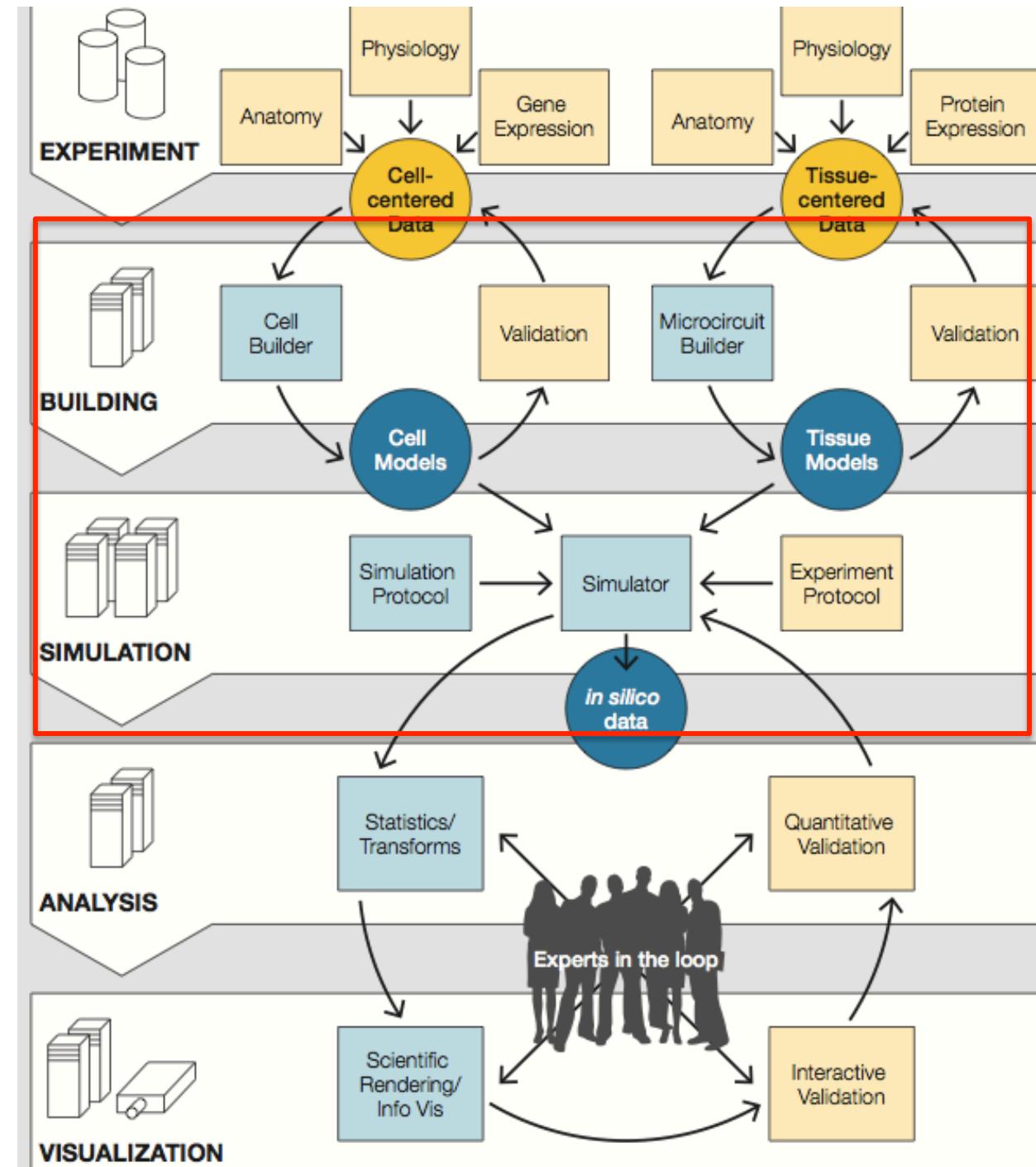
- Migliore et al, 2006
- Druckmann et al, 2007
- Druckmann et al, 2008
- Hines et al, 2008 a & b
- Kozloski et al, 2008
- Hay et al, 2011
- Hines et al, 2011
- Ranjan et al, 2011
- Lasserre et al, 2012
- Druckmann et al, 2012
- Hill et al, 2012
- Tauheed et al, 2012
- Hernando et al, 2012
- Ramaswamy et al, 2012
- Reimann et al, 2013
- Schürmann et al, 2014
- Delalondre et al, 2014
- Devresse et al, 2015
- Kumbhar et al, 2016
- ...



Co-designed Heterogeneous ICT Facility

Relevant publications:

- Migliore et al, 2006
- Druckmann et al, 2007
- Druckmann et al, 2008
- Hines et al, 2008 a & b
- Kozloski et al, 2008
- Hay et al, 2011
- Hines et al, 2011
- Ranjan et al, 2011
- Lasserre et al, 2012
- Druckmann et al, 2012
- Hill et al, 2012
- Tauheed et al, 2012
- Hernando et al, 2012
- Ramaswamy et al, 2012
- Reimann et al, 2013
- Schürmann et al, 2014
- Delalondre et al, 2014
- Devresse et al, 2015
- Kumbhar et al, 2016
- ...



(European) Human Brain Project

- Building infrastructure dedicated to Neuroscience
- EU Flagship project initiated by EPFL/BBP gathering more than 100 universities/labs
- Started in September 2013 with budget/proposal submission every 2 years (Total budget expected over 10 years ~ 1 billion euros)

Building Infrastructure Dedicated to Neuroscience

PaaS/IaaS

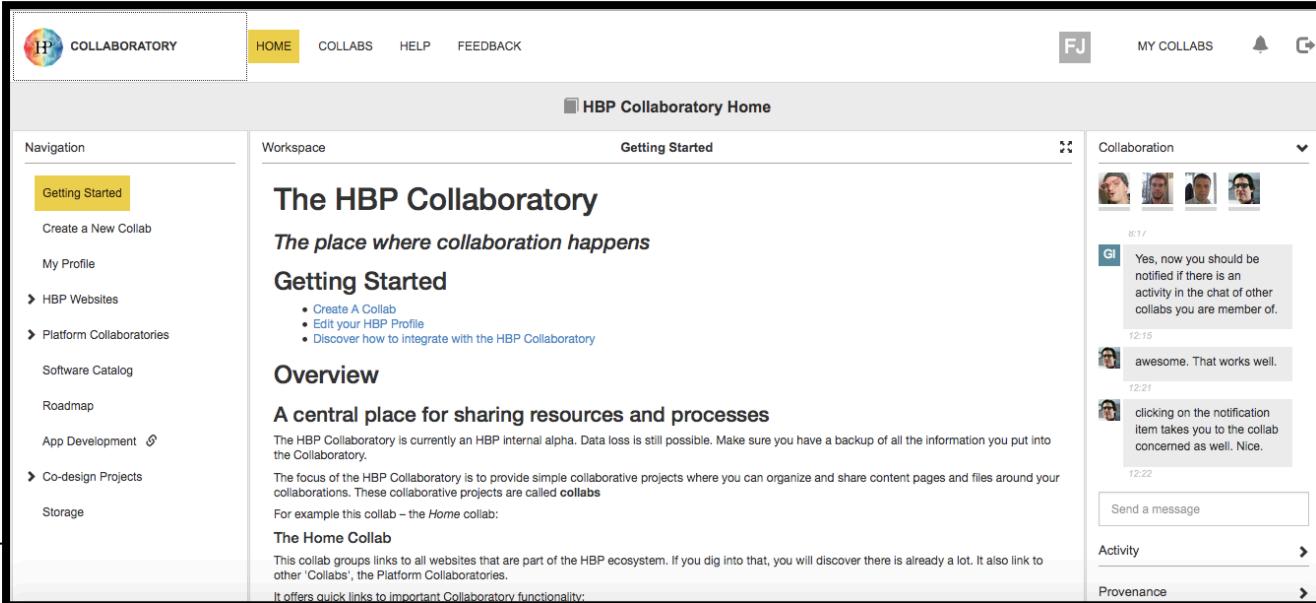
- Neuroinformatics
- Brain Simulation
- High Performance Analytics and Computing
- Medical informatics
- Neuromorphic Computing
- Neurorobotics

Building Infrastructure Dedicated to Neuroscience

PaaS/IaaS

- Neuroinformatics
- Brain Simulation
- High Performance Analytics and Computing
- Medical informatics
- Neuromorphic Computing
- Neurorobotics

**Accessible through
HBP Web-based
Collaboratory
(SaaS)**



The screenshot shows the HBP Collaboratory Home page. The left sidebar contains a navigation menu with links to 'Getting Started' (which is highlighted), 'Create a New Collab', 'My Profile', 'HBP Websites', 'Platform Collaboratories', 'Software Catalog', 'Roadmap', 'App Development', 'Co-design Projects', and 'Storage'. The main workspace displays the 'The HBP Collaboratory' section, which includes the subtitle 'The place where collaboration happens', a 'Getting Started' section with three bullet points, and an 'Overview' section with a sub-section 'A central place for sharing resources and processes'. On the right side, there is a 'Collaboration' panel showing a list of users and their recent messages. One message from user 'GJ' says: 'Yes, now you should be notified if there is an activity in the chat of other collabs you are member of.' Another message from user 'FJ' says: 'awesome. That works well.' A third message from user 'GJ' says: 'clicking on the notification item takes you to the collab concerned as well. Nice.' Below the collaboration panel, there are sections for 'Activity' and 'Provenance'.

Building Infrastructure Dedicated to Neuroscience

PaaS/IaaS

- Neuroinformatics
- Brain Simulation
- High Performance Analytics and Computing
- Medical informatics
- Neuromorphic Computing
- Neurorobotics

**Accessible through
HBP Web-based
Collaboratory
(SaaS)**



- Blue Brain Project (BBP) & Human Brain Project (BBP) introduction
- **On node portable performance optimization**
- Future R&D Directions

Portable On Node Optimization

- **What scientific problem we are trying to solve ?**
- What is our development workflow/Tools ?

Biological Problem to Solve

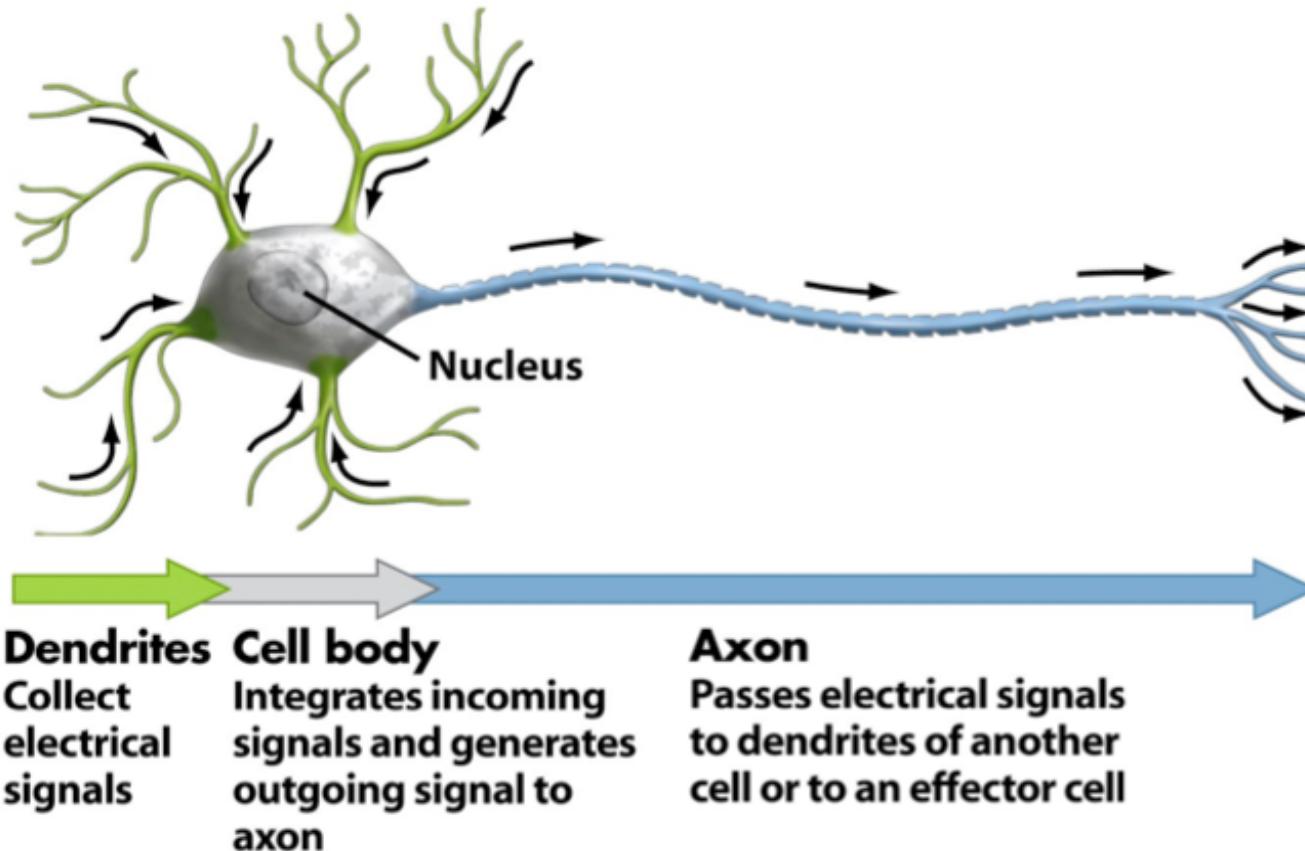
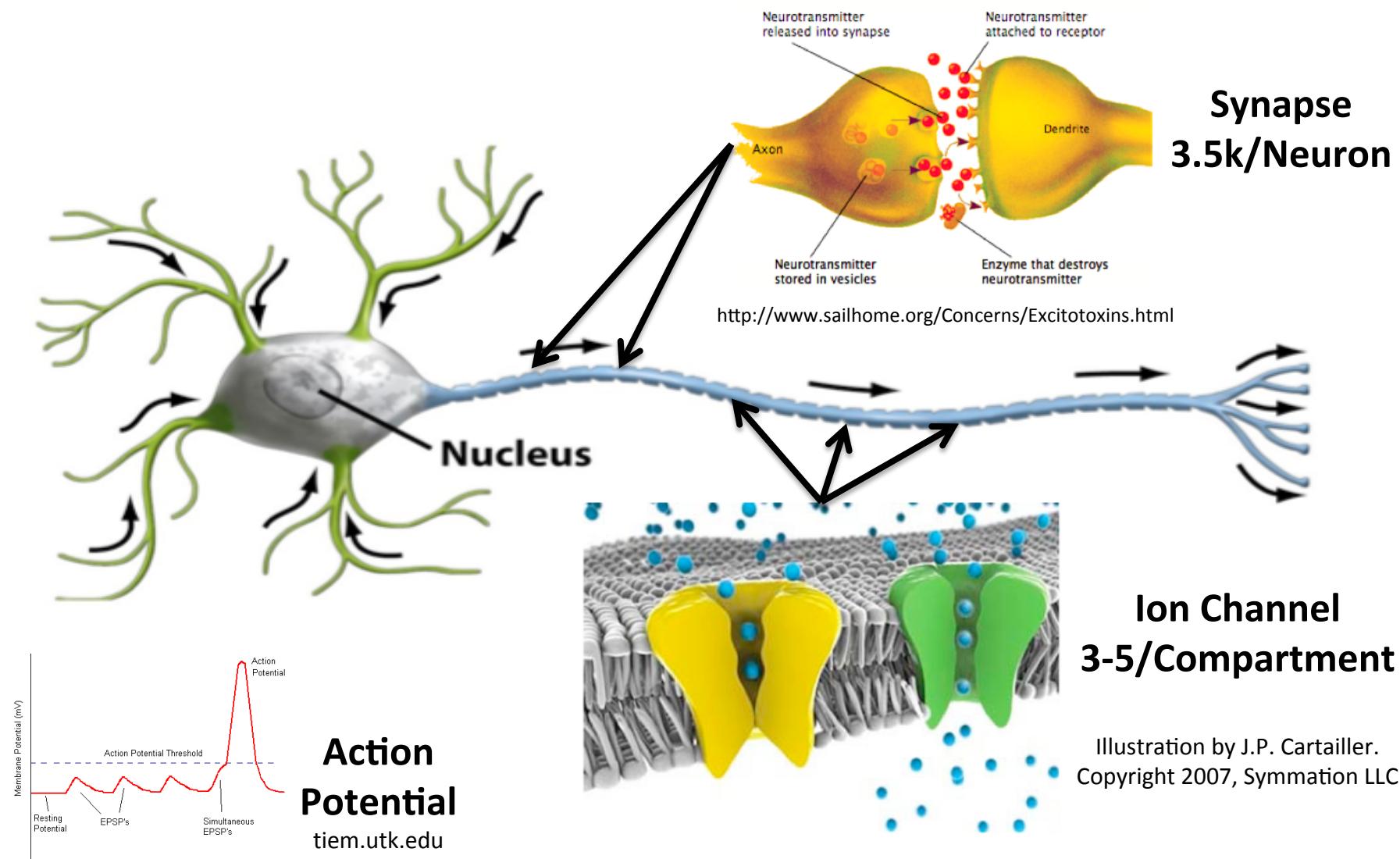
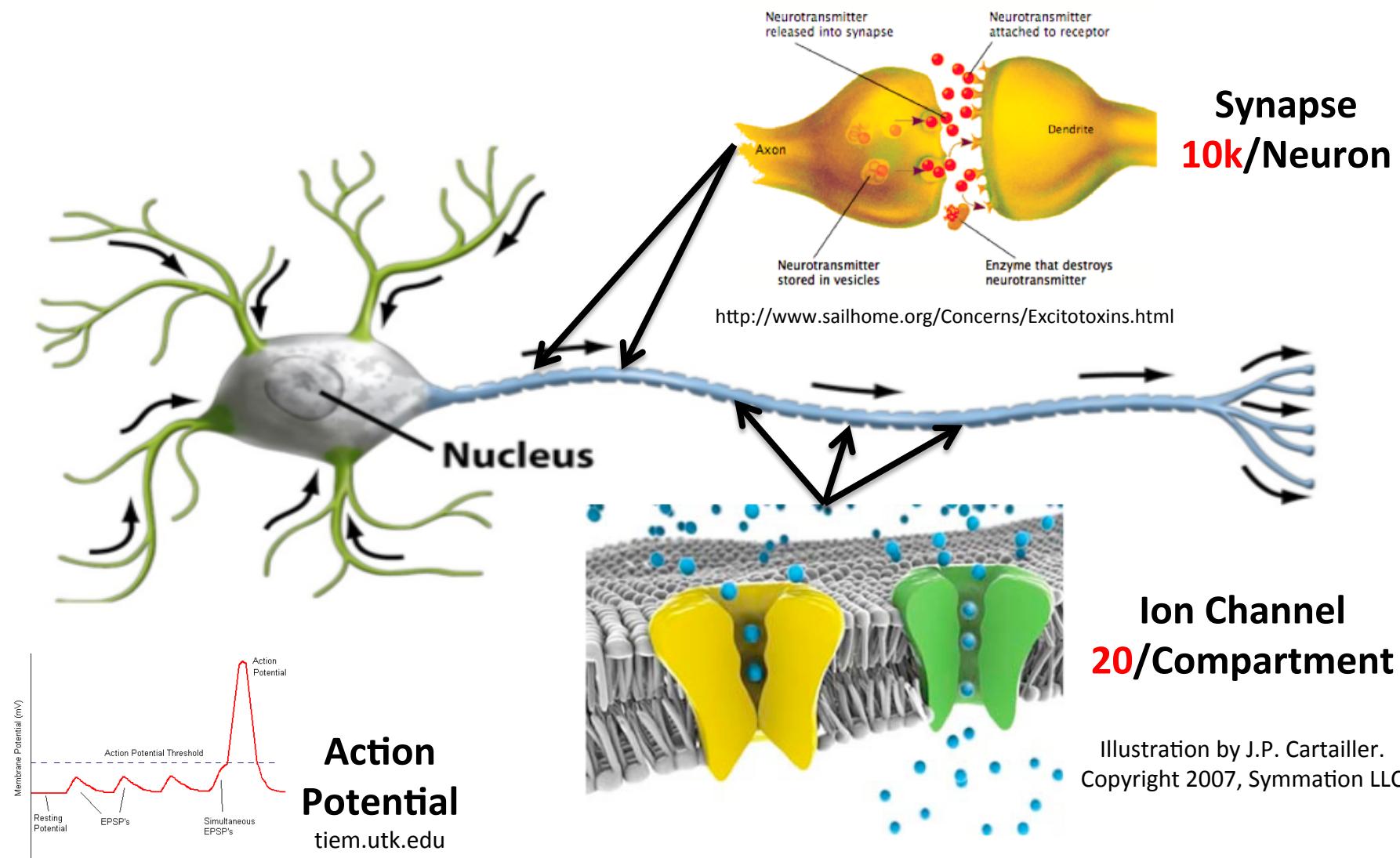


Figure 45-2b Biological Science, 2/e
© 2005 Pearson Prentice Hall, Inc.

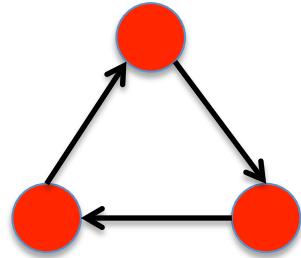
Biological Problem to Solve Today



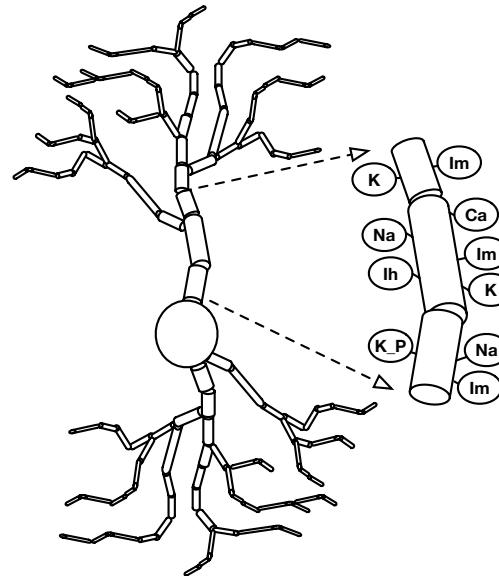
Biological Problem to Solve Tomorrow



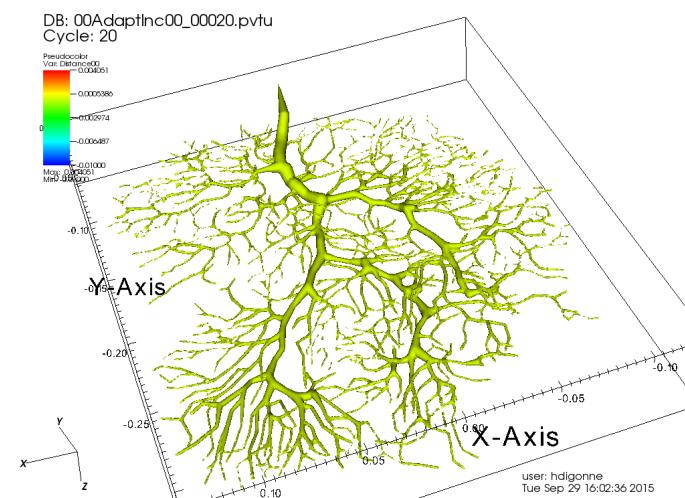
Different Scales/Different Representations



**Point Neuron
(NEST)**

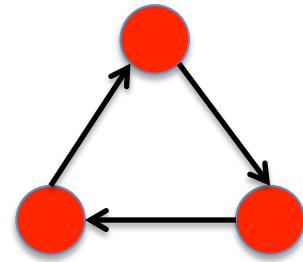


**Morphologically Detailed
(NEURON/CoreNeuron)**

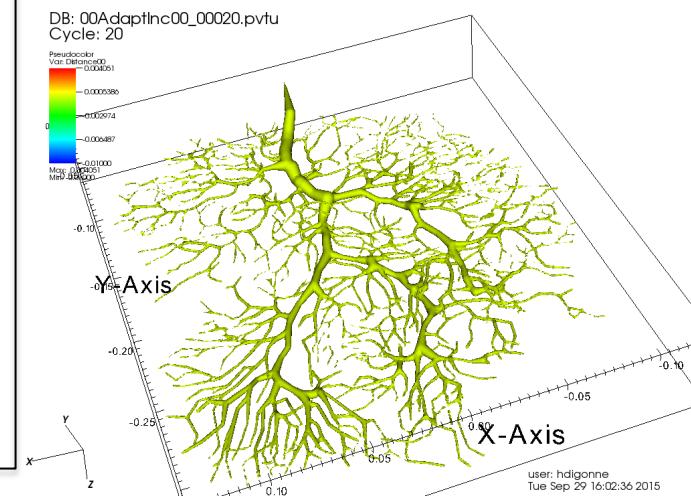
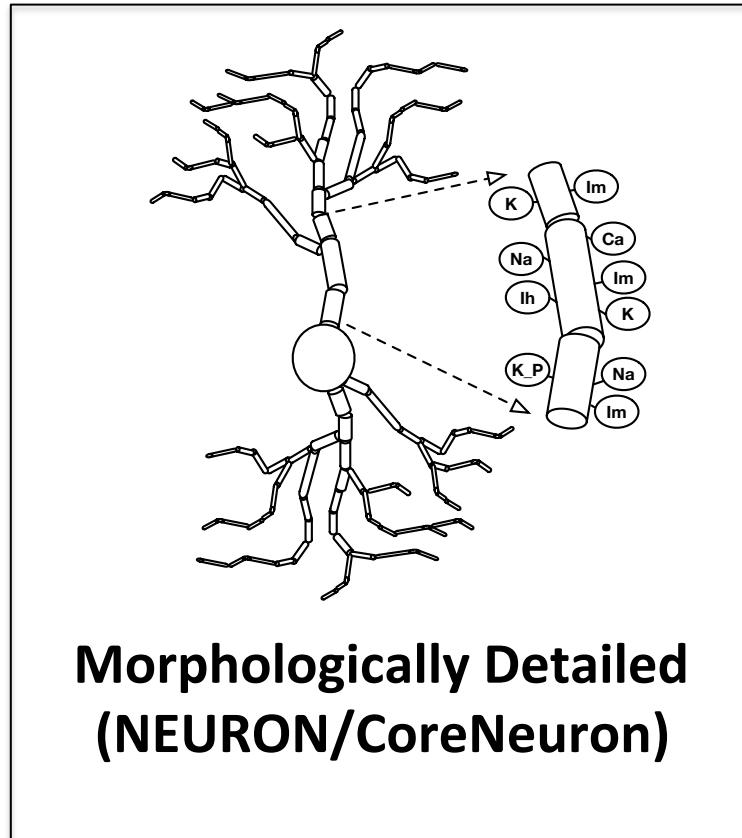


**3D Modeling
(STEPS)**

Different Scales/Different Representations

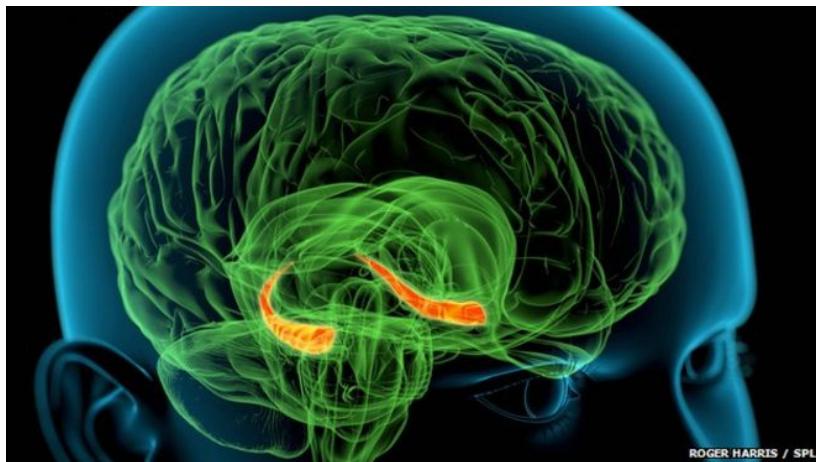


**Point Neuron
(NEST)**



**3D Modeling
(STEPS)**

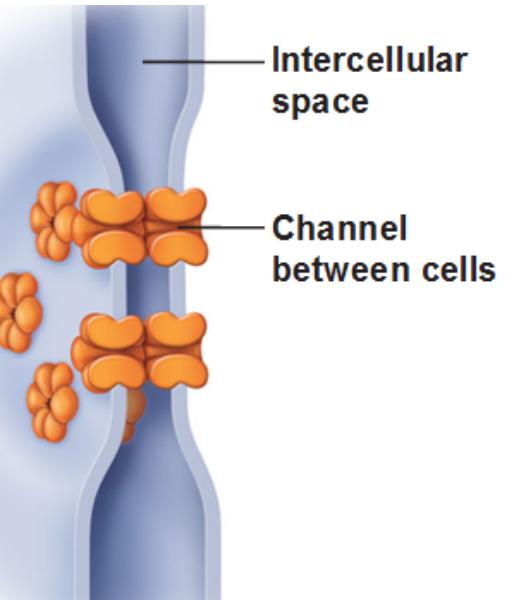
Subset of Use Cases with Very Diverse Requirements



Hippocampus
Armando Romani [EPFL]

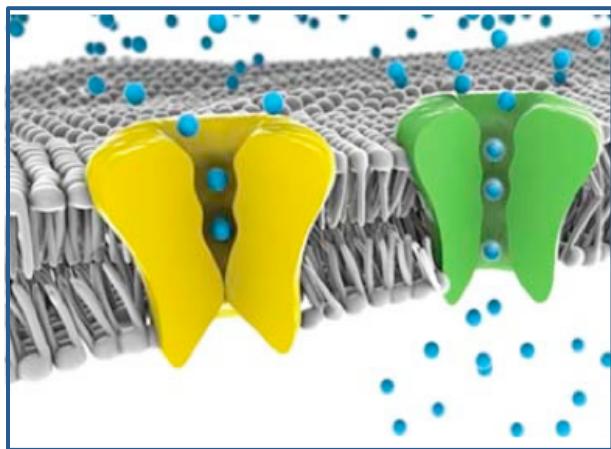


Structural Plasticity
Giuseppe Chindemi [EPFL]

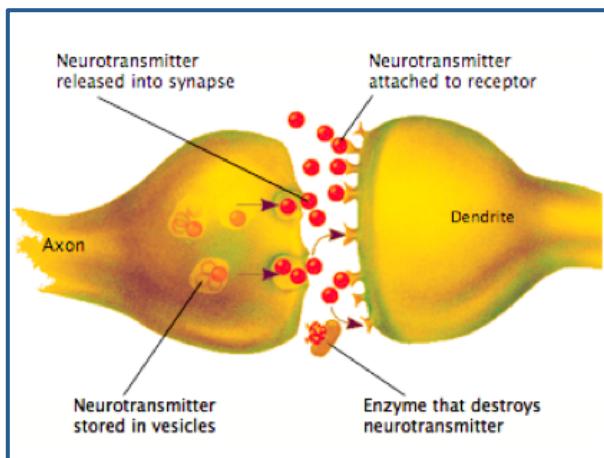


Gap Junction
Oren Amsalem [HUJI]

Resolution Workflow



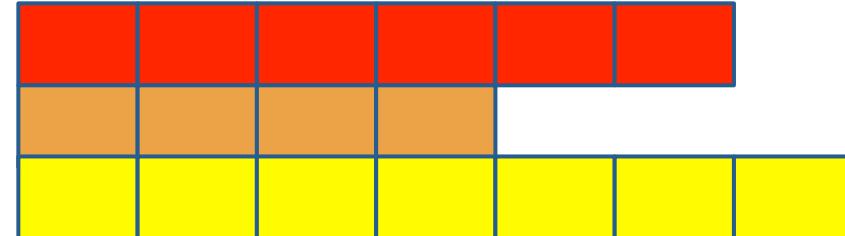
Ion Channel



Synapse

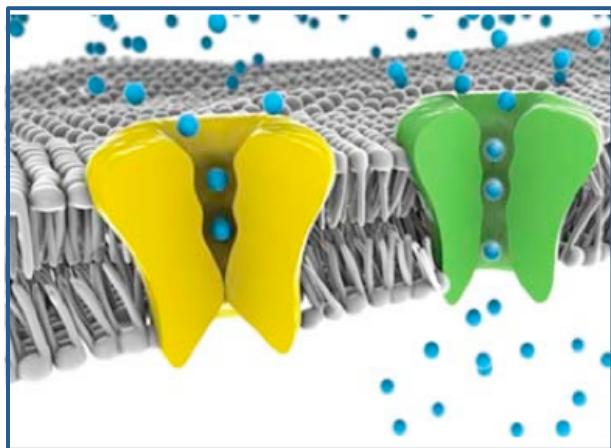
Mechanism Type

Mechanism Instance

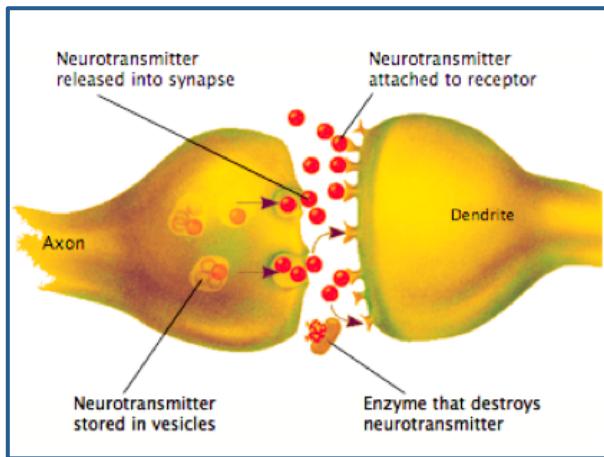


Data Management Organization

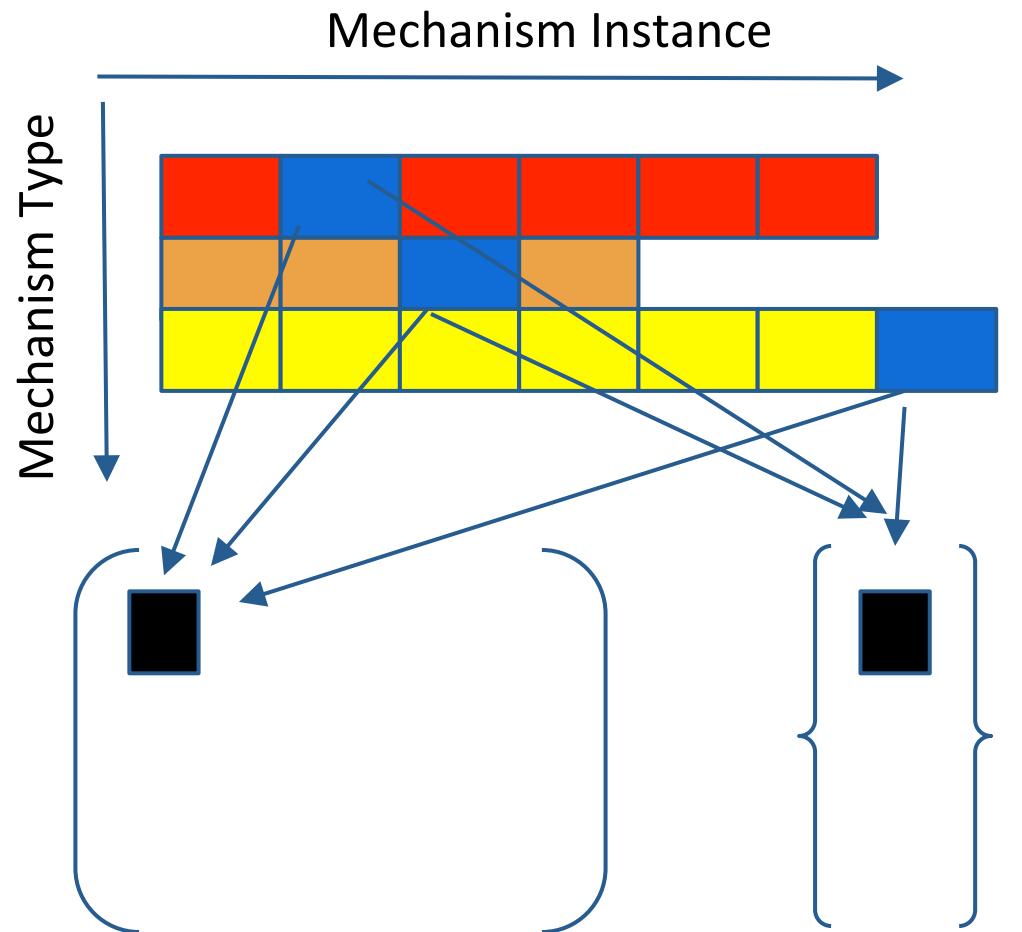
Resolution Workflow



Ion Channel



Synapse



Kernels – 85% of Total Compute Time

Compute Bound – 20% of Kernels

```
/* loop2: this is second kernel with more compute due to exp and div */
__PRAGMA_FOR_VECTOR_LOOP_
for( i = 0; i < nodecount; i++) {

    int idx = nt->node_index[i];
    double v = nt->vec_v[idx];

    p3[i] = data[ion1[i]];

    double qt = 2.952882641412121;

    double mAlpha = (0.182*(v+32.0)) / (1.0-(exp(-v-32.0)/6.0));
    double mBeta = (0.124*(-v-32.0)) / (1.0-(exp(v+32.0)/6.0));

    double mInf = mAlpha/(mAlpha+mBeta);
    double mTau = (1.0/(mAlpha+mBeta))/qt;

    p1[i] = p1[i] + (1.0-exp(dt*(-1.0/mTau))) * (-(mInf/mTau)/ ((-1.0/mTau)-p1[i]));

    double hAlpha = (-0.015*(v+60.0)) / (1.0-(exp(( v+60.0)/6.0)));
    double hBeta = (-0.015*(-v-60.0)) / (1.0-(exp((-v-60.0)/6.0)));
    double hInf = hAlpha / (hAlpha+hBeta);
    double hTau = (1.0/(hAlpha+hBeta)) / qt;

    p2[i] = p2[i] + (1.0-exp(dt*(-1.0/hTau))) * (-(hInf/hTau)/ ((-1.0/hTau)
nt->vec_v[idx] = v;

    p5[i] += v;
}
```

Memory Bound – 80% of Kernels

```
/* loop1: this is one kernel with less compute / memory streaming */
__PRAGMA_FOR_VECTOR_LOOP_
for( i = 0; i < nodecount; i++) {

    int idx = nt->node_index[i];
    double v = nt->vec_v[idx];

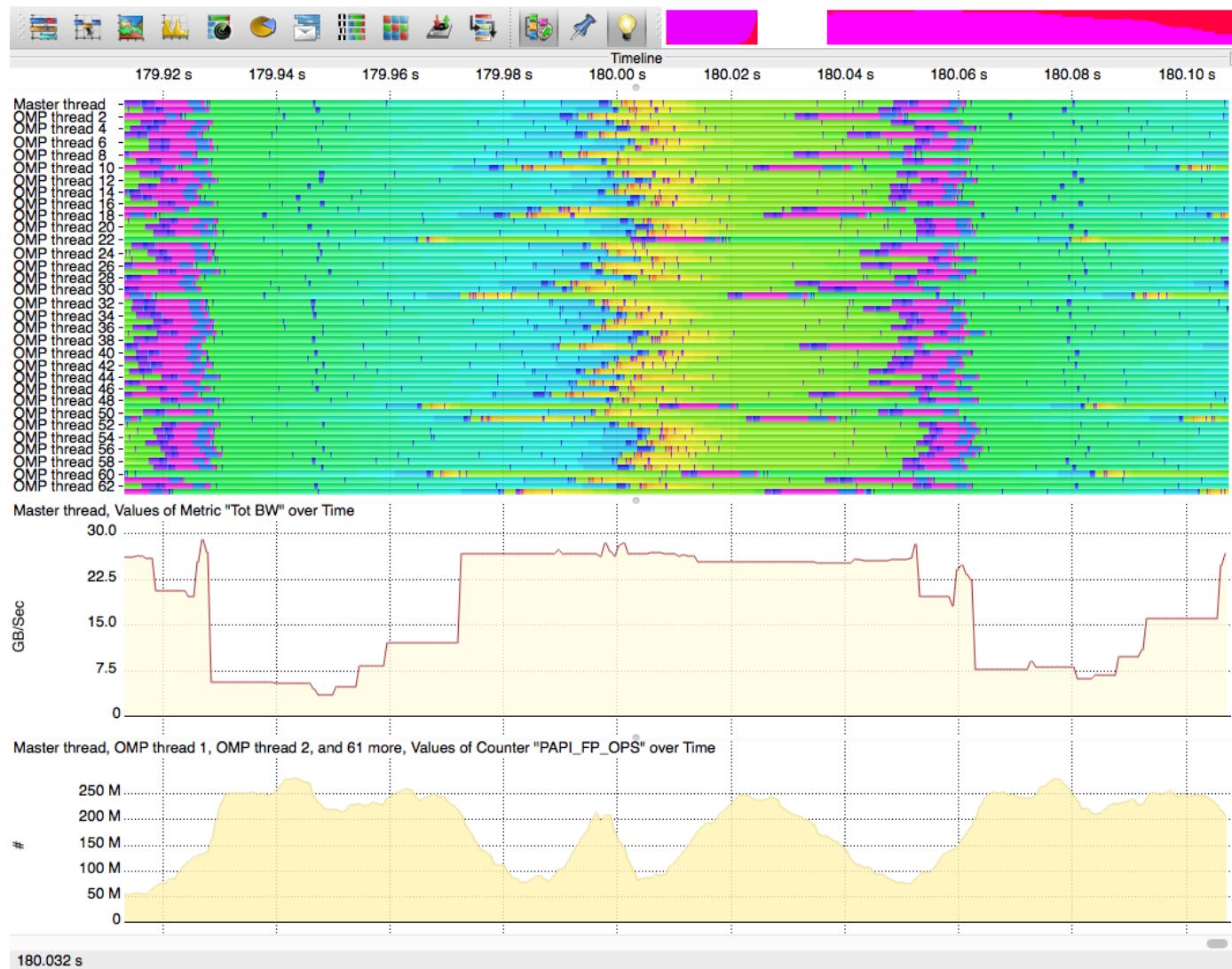
    p3[i] = data[ion1[i]];

    double gNaTs2 = p1[i] * p2[i] * p2[i] * p2[i] + p3[i];
    double ina = gNaTs2 * ( v - p3[i] );

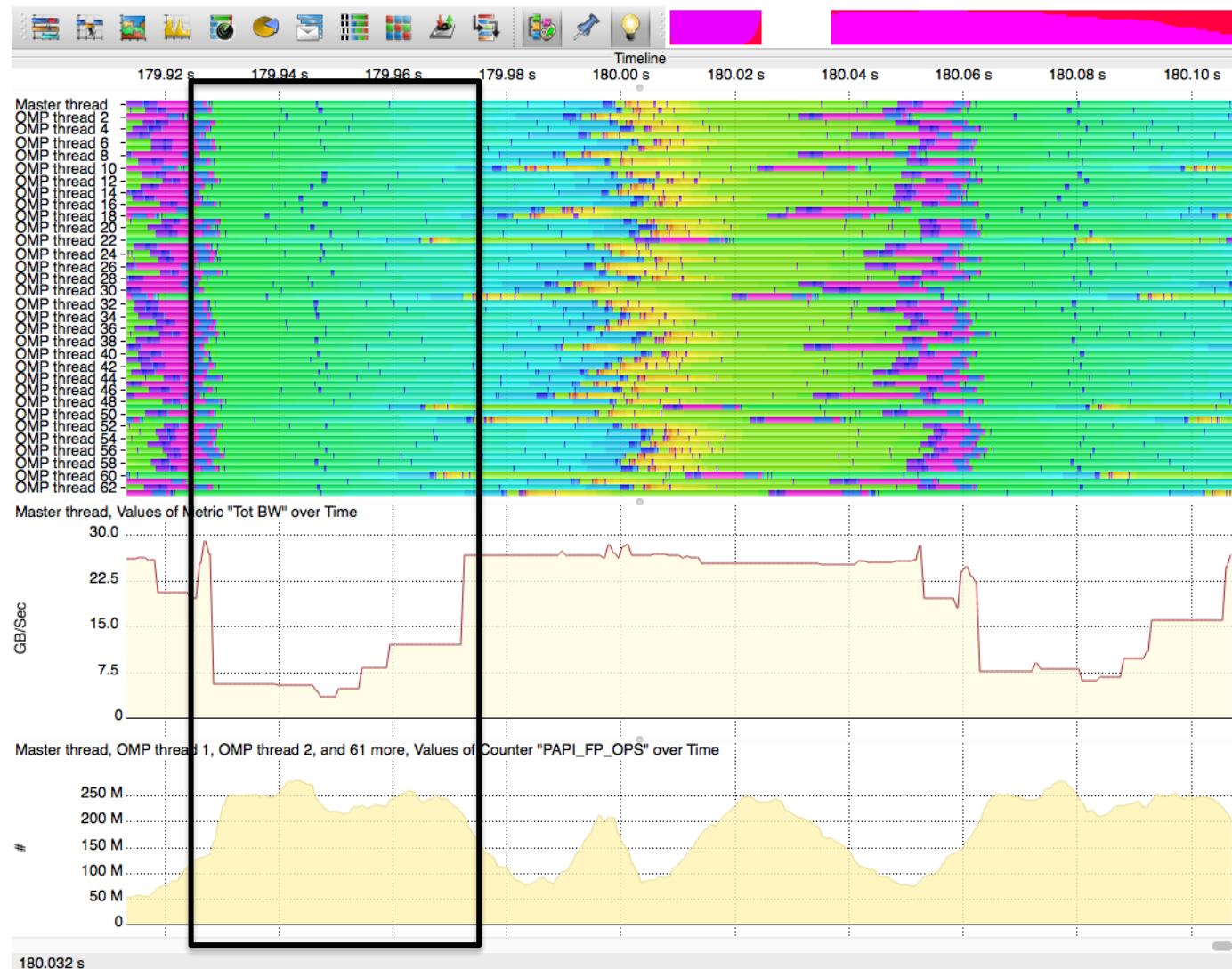
    /* no data hazard below */
    data[ion1[i]] += gNaTs2;
    data[ion2[i]] += ina;

    nt->vec_rhs[idx] -= ina;
    nt->vec_d[idx] += gNaTs2;
}
```

Compute/Bandwidth Analysis

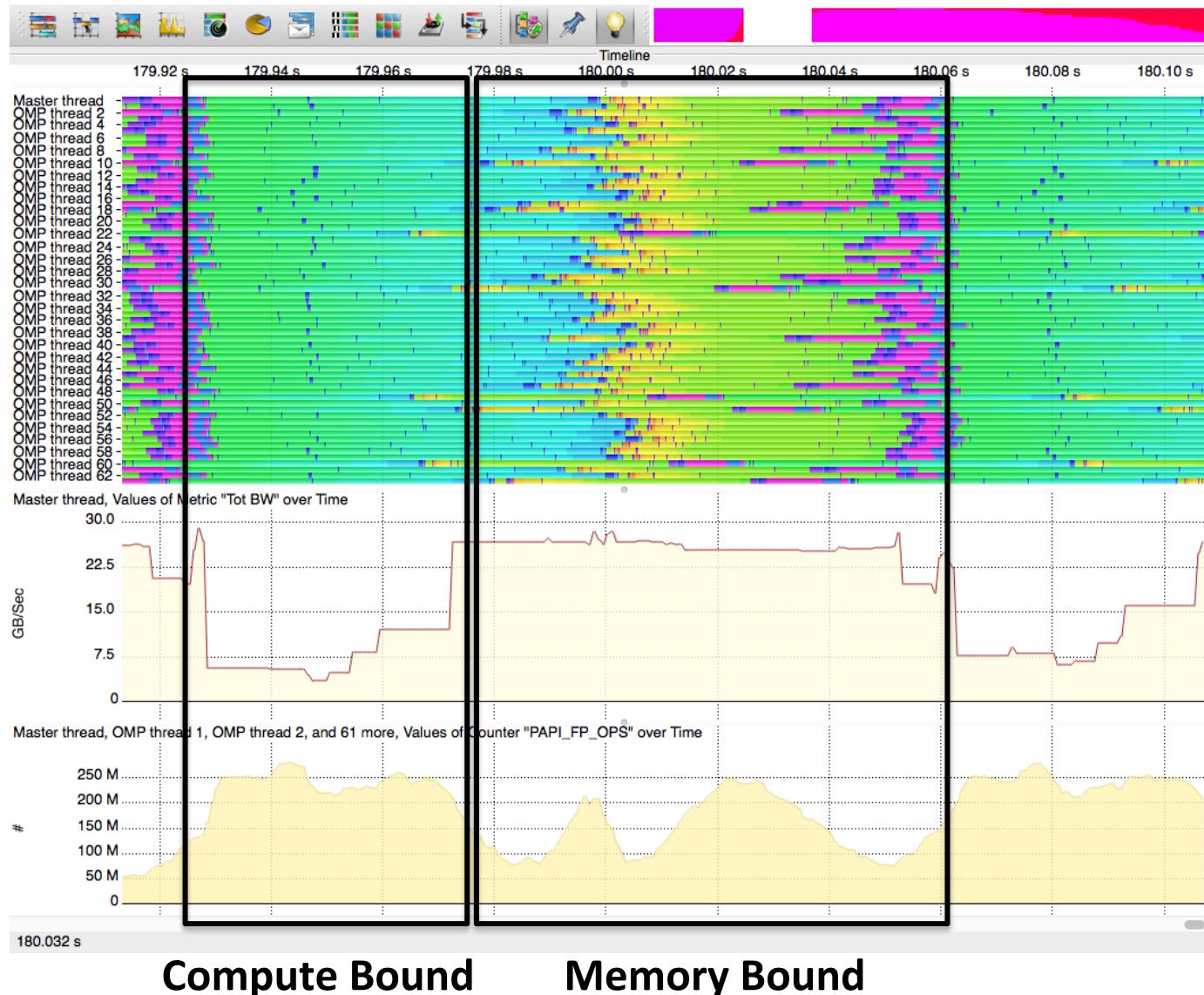


Compute/Bandwidth Analysis



Compute Bound

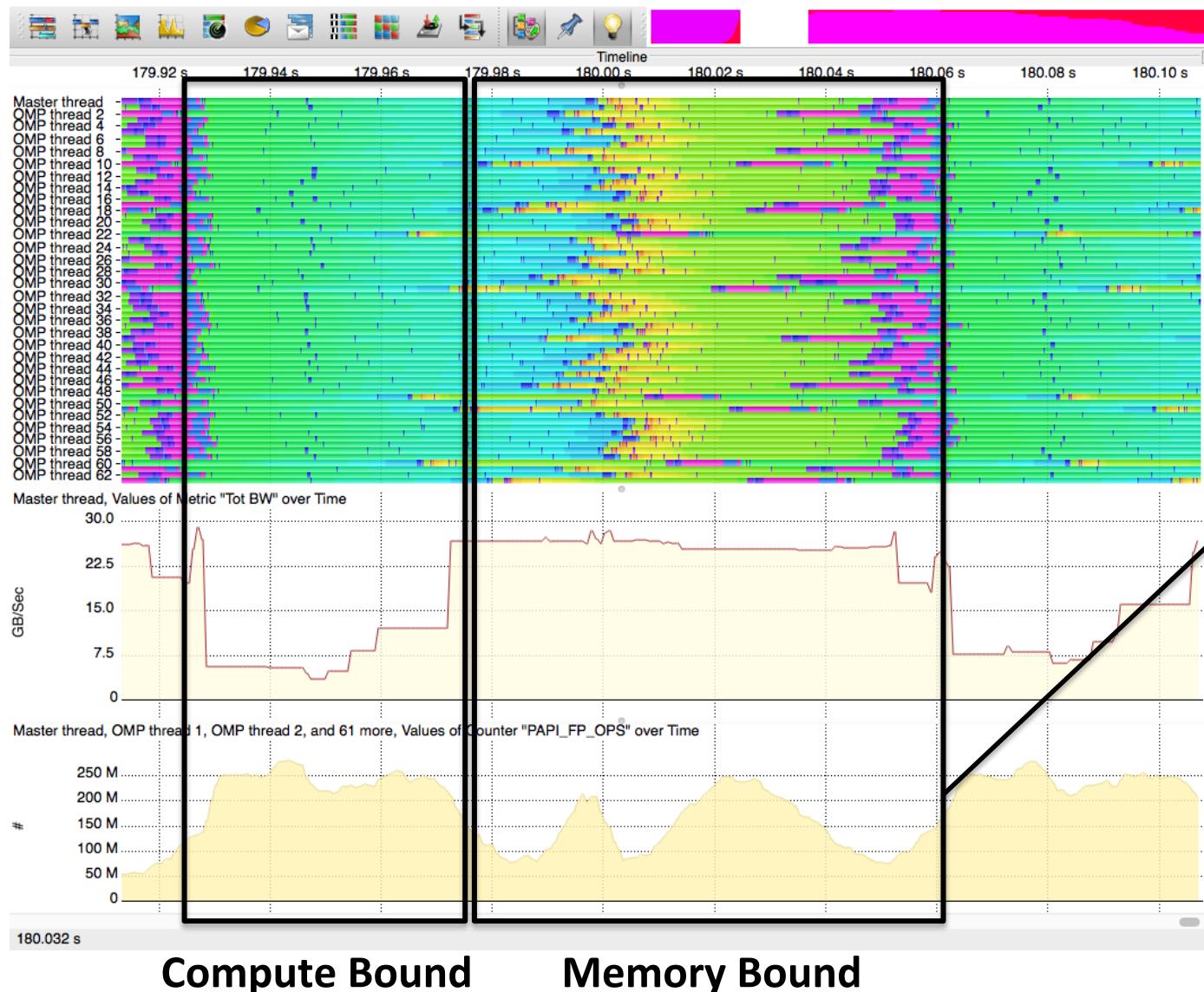
Compute/Bandwidth Analysis



Compute Bound

Memory Bound

Compute/Bandwidth Analysis



Looking at ways to
overlap kernels

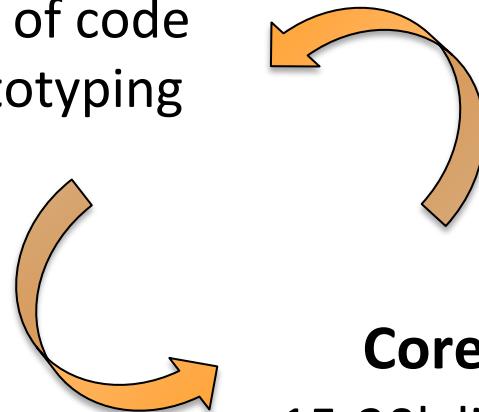
Portable On Node Optimization

- What scientific problem we are trying to solve ?
- What is our development workflow/Tools ?

Development Workflow & Tools

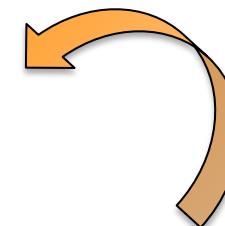
NeuroM(ini)app

150-200 lines of code
Used for prototyping



CoreNeuron

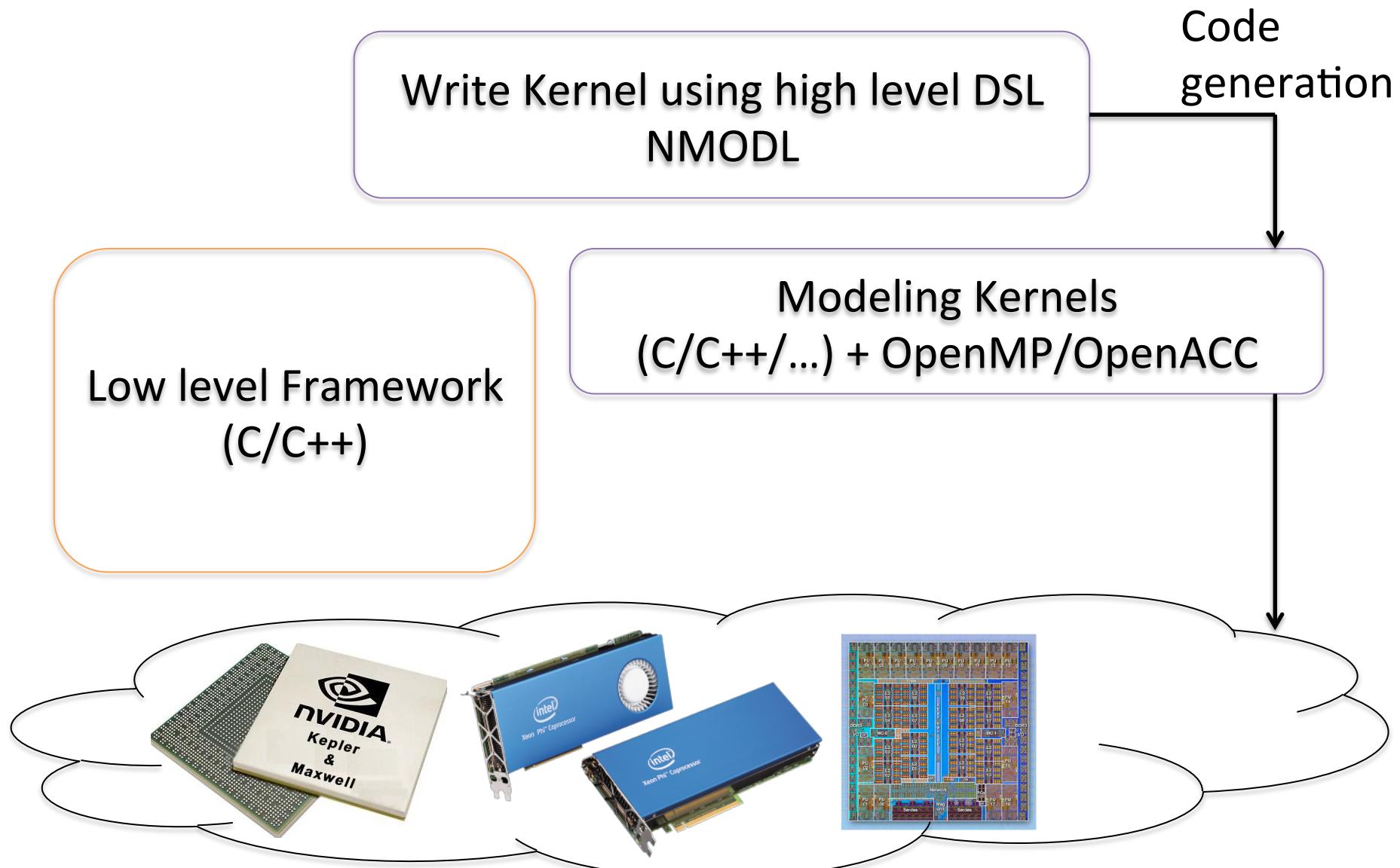
15-20k lines of code
Optimized kernels



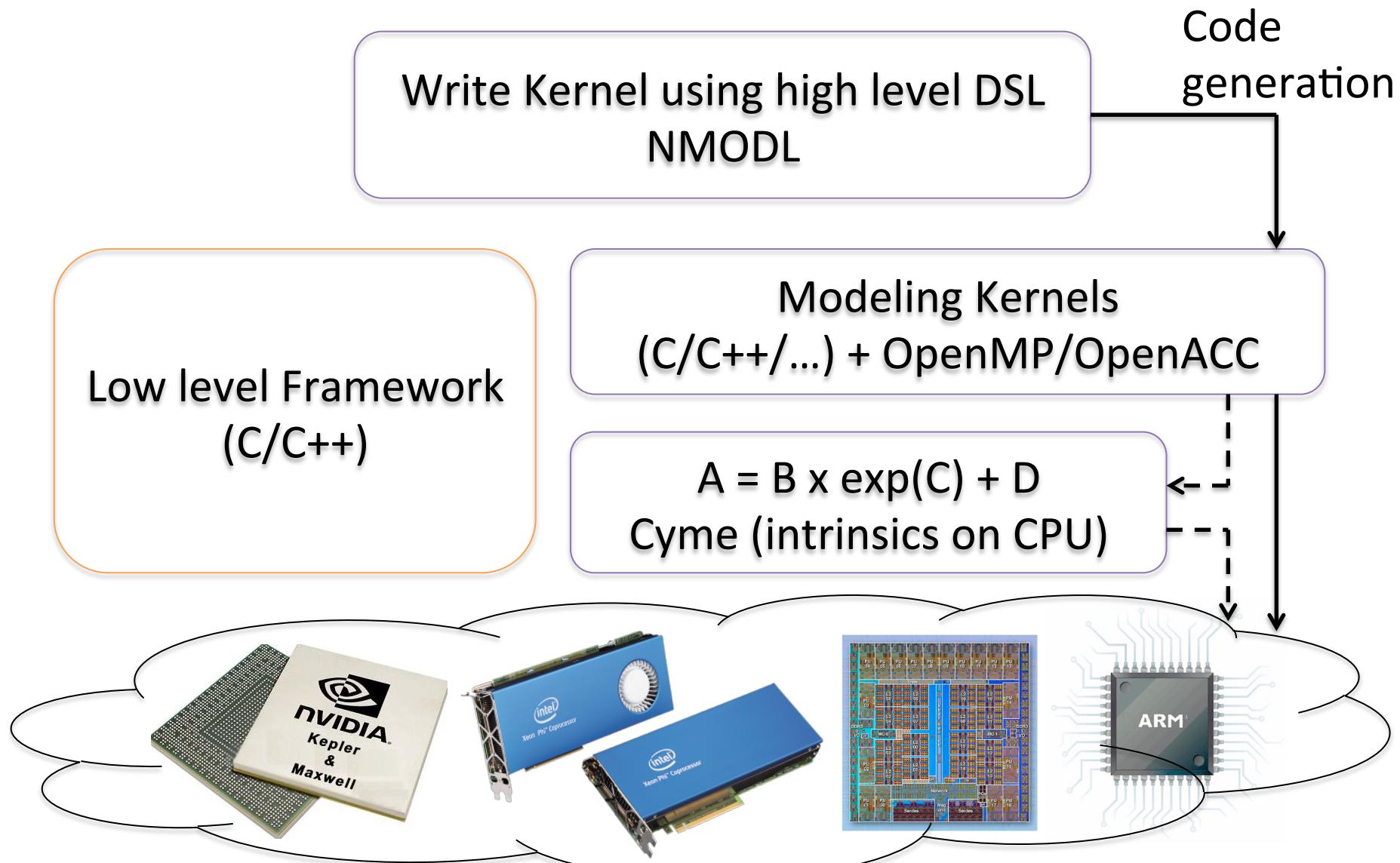
NEURON

200k lines of code
All functionalities

Performance Portability Through DSL



Performance Portability Through DSL



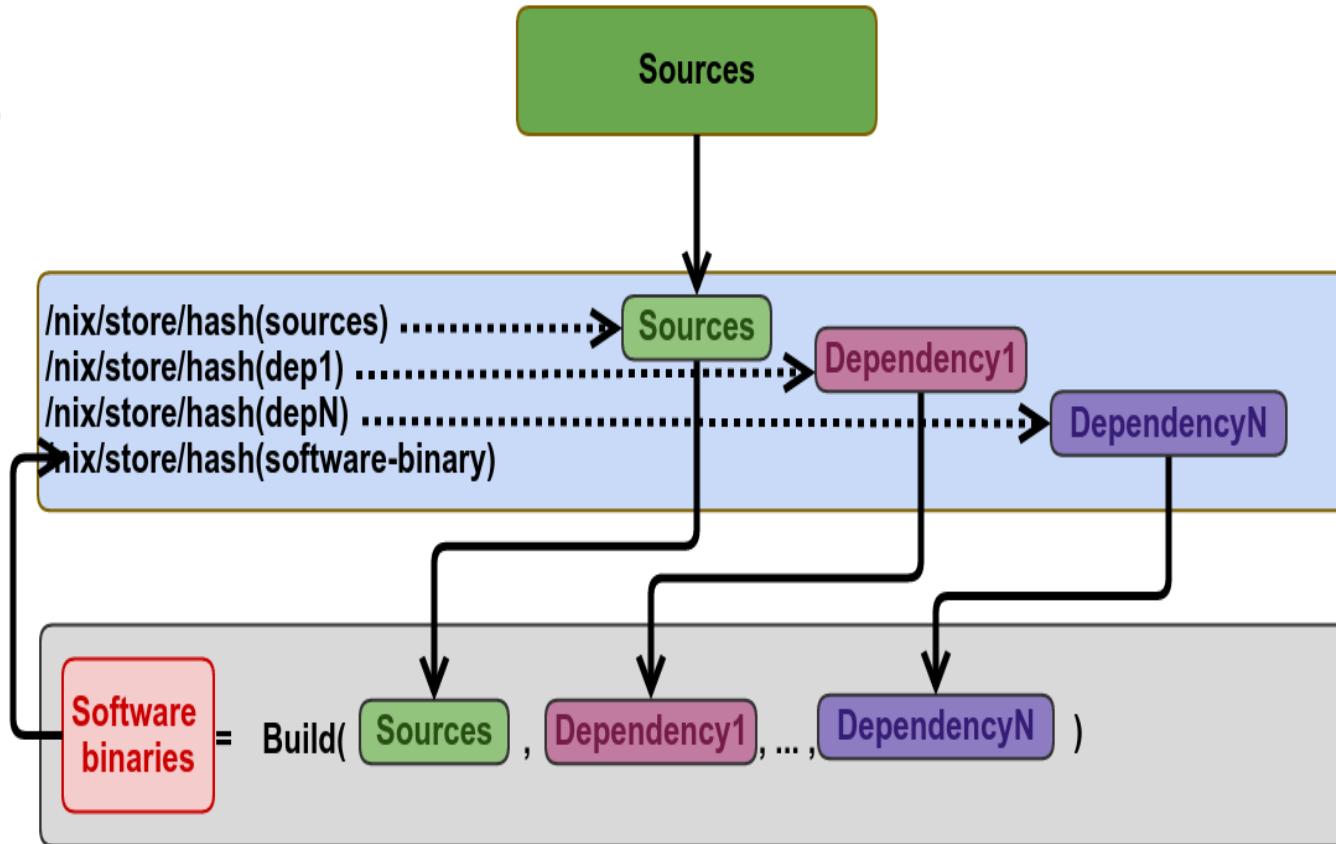
Software Portability & Reproducibility

- Scientific reproducibility **is a major issue !**
- Modules or equivalent too fragile/unpredictable
- RPM/Debian packages update break development environment
- Docker/Containers come with problems today...

**Source
repo (VCS)**

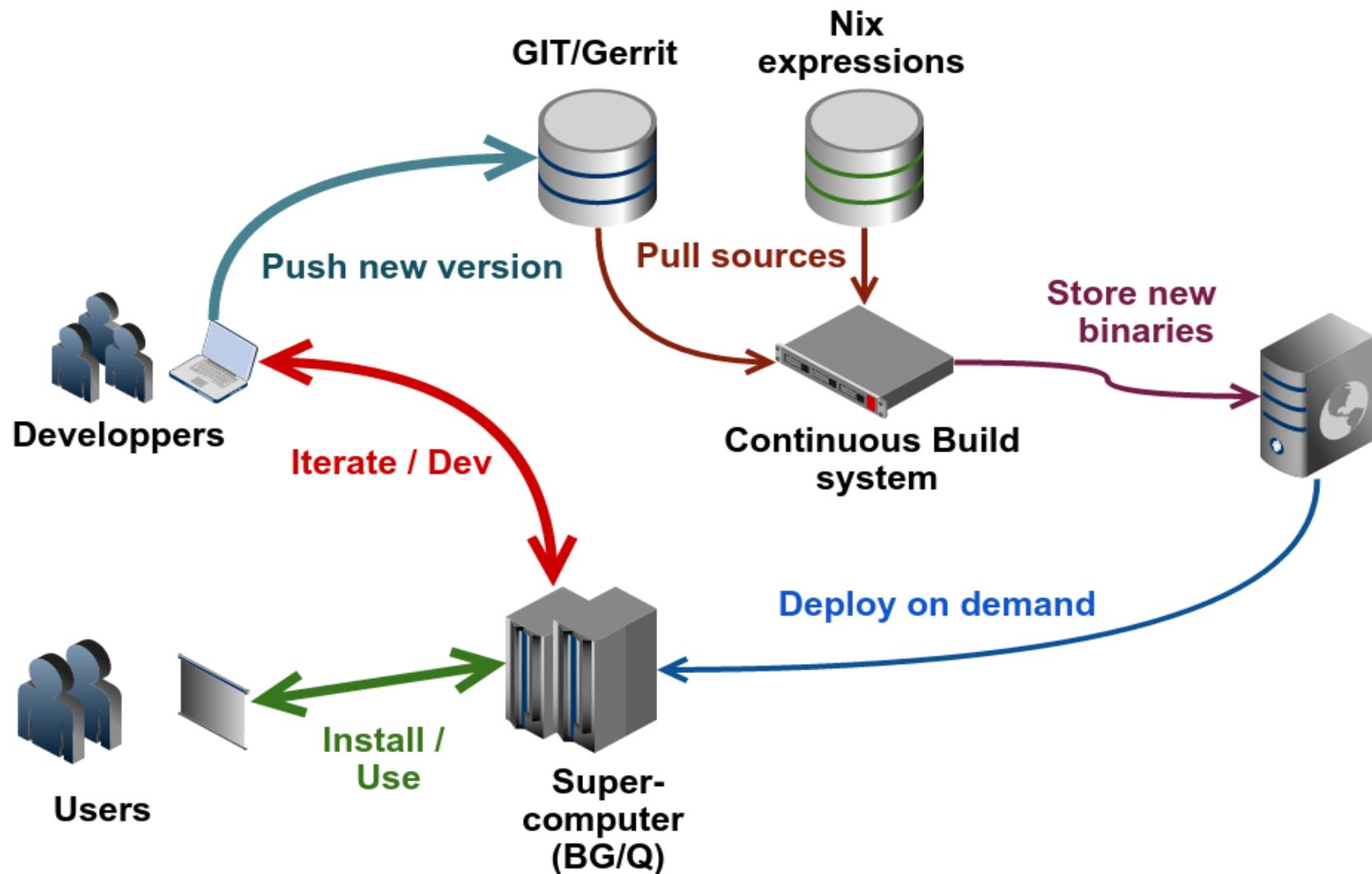
Nix store

Nix build



More information: A. Devresse et al., Fully Automated Workflows and Ecosystem to guarantee Scientific Result Reproducibility across Platforms, Software Environment and Systems, SC15 Conference

Continuous Integration & Deployment



Application Status

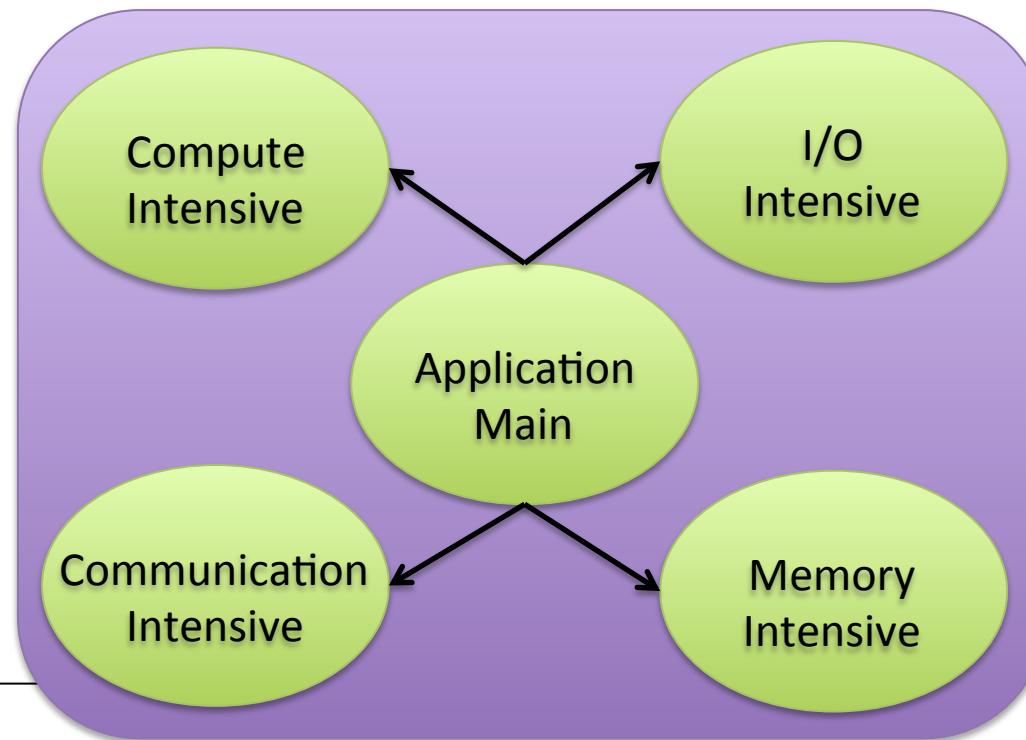
- Intel x86, KNC, Nvidia GPU, IBM Blue Gene/Q & ARM supported
- Weak scaling to full MIRA with less than 20% loss of efficiency
- All kernel performance profiled/classified
- Nix & Continuous integration/deployment moving to production (installation at BBP & JSC)
- New challenging use cases ... (Real time, ...)

- Blue Brain Project (BBP) & Human Brain Project (BBP) introduction
- On node portable performance optimization
- **Future R&D Directions**

- Co-design Heterogeneous Infrastructure
- Development Workflow
- Leveraging Deep Memory Hierarchy
- Scientific Workflows

Open Questions

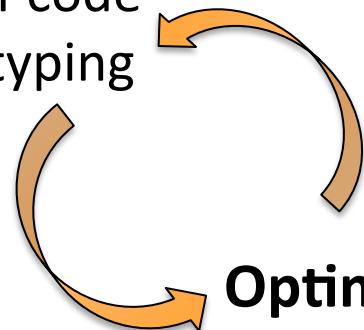
- Is this a system node or an eco-system ?
- How can I best map kernels to hardware component ?
- How can I explore workflow opportunities ?
- How can I build a system which fits my requirements ?
- How can I continuously collect application requirements ?



NeuroM(ini)app

150-200 lines of code

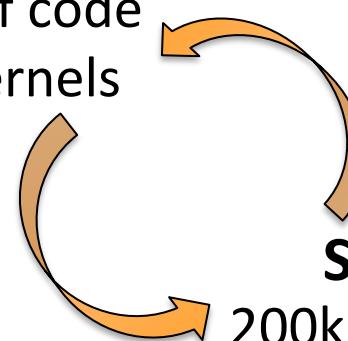
Used for prototyping



Optimized App.

15-20k lines of code

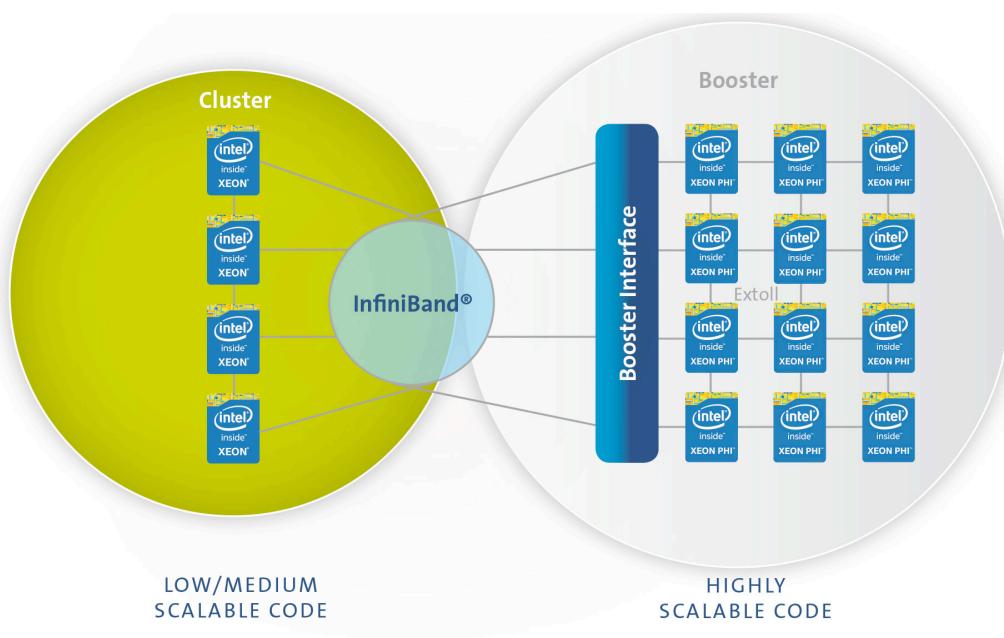
Optimized kernels



Sc. App.

200k lines of code

All functionalities



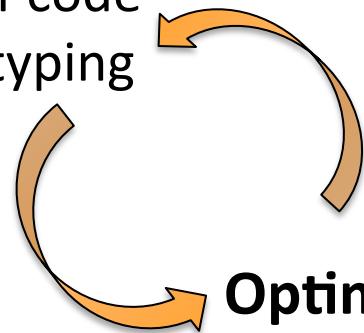
**Dynamic Exascale Entry Platform
(DEEP)**

**HBP Precommercial
Procurement (PCP)**

NeuroM(ini)app

150-200 lines of code

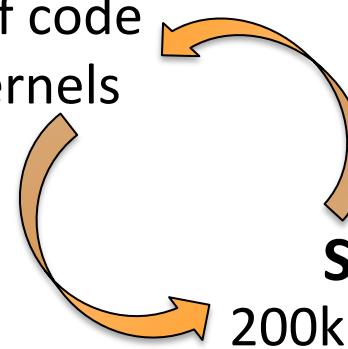
Used for prototyping



Optimized App.

15-20k lines of code

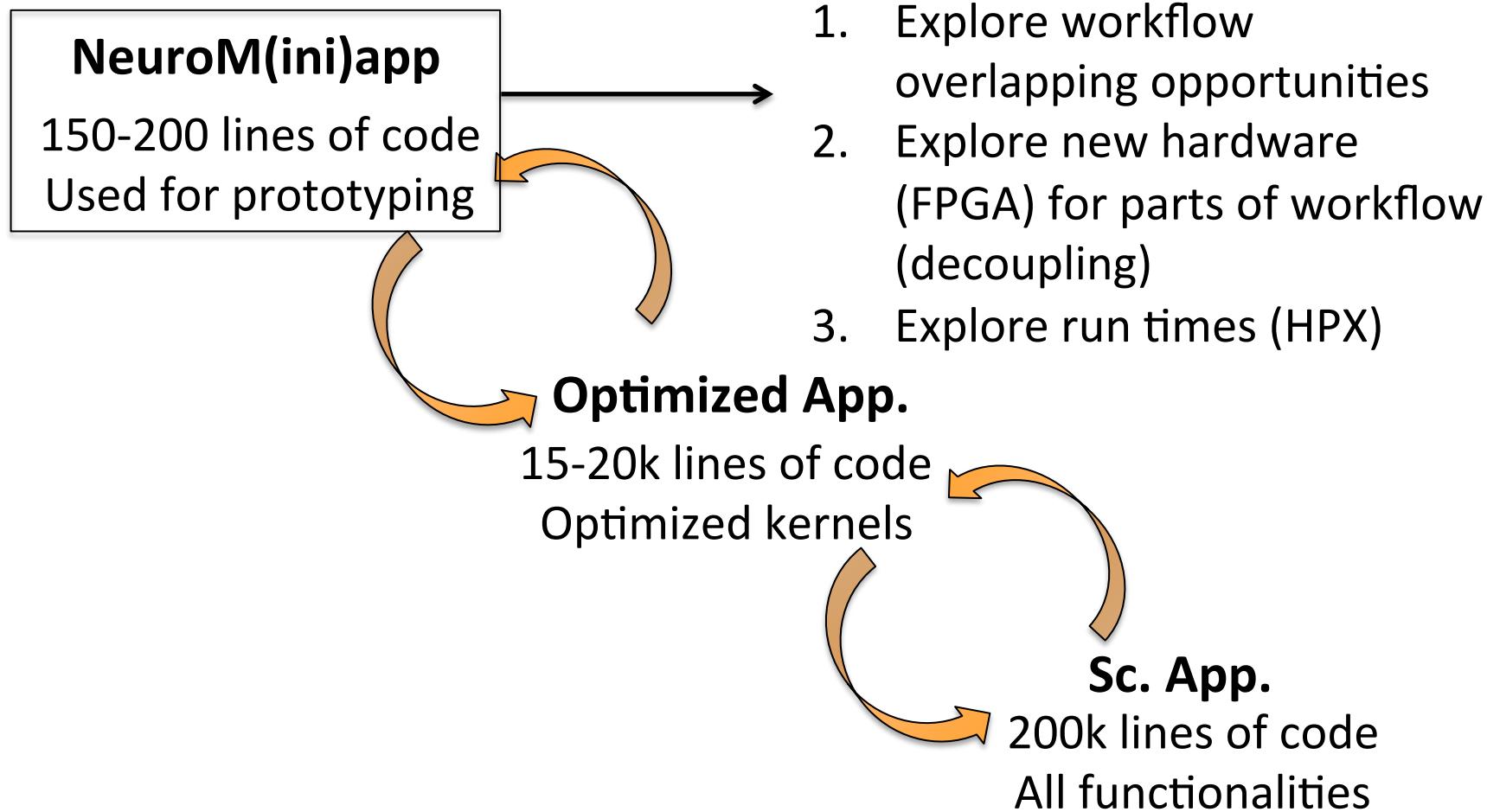
Optimized kernels



Sc. App.

200k lines of code

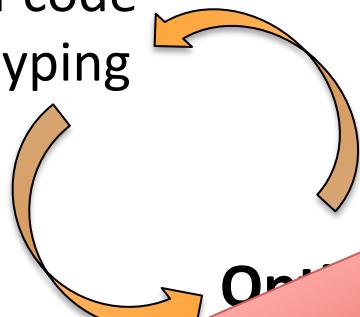
All functionalities



NeuroM(ini)app

150-200 lines of code

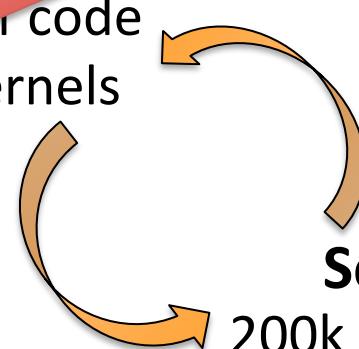
Used for prototyping



OpenM

Good ... but very time consuming/Error-prone !!

100k lines of code
Optimized kernels



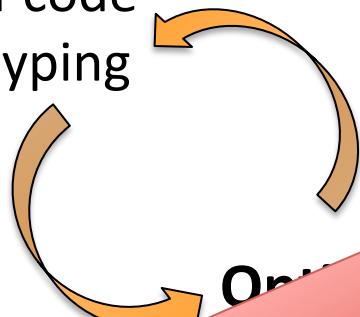
Sc. App.

200k lines of code
All functionalities

NeuroM(ini)app

150-200 lines of code

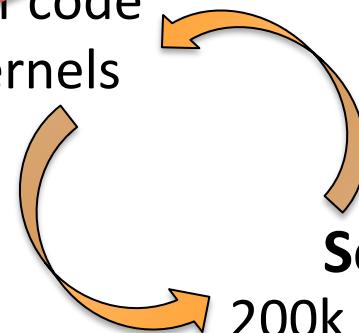
Used for prototyping



Open-M

Good ... but very time consuming/Error-prone !!

100k lines of code
Optimized kernels

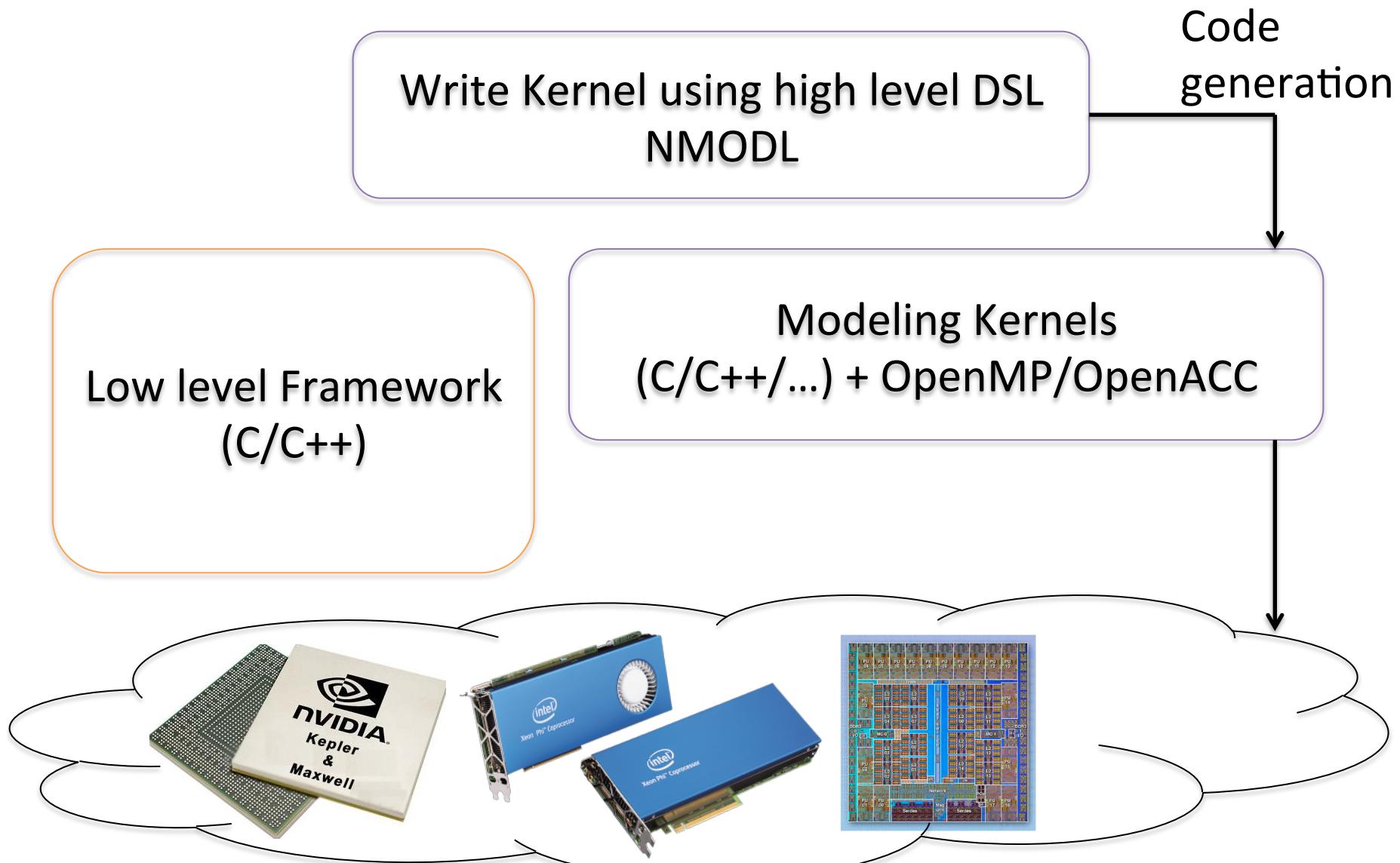


Sc. App.

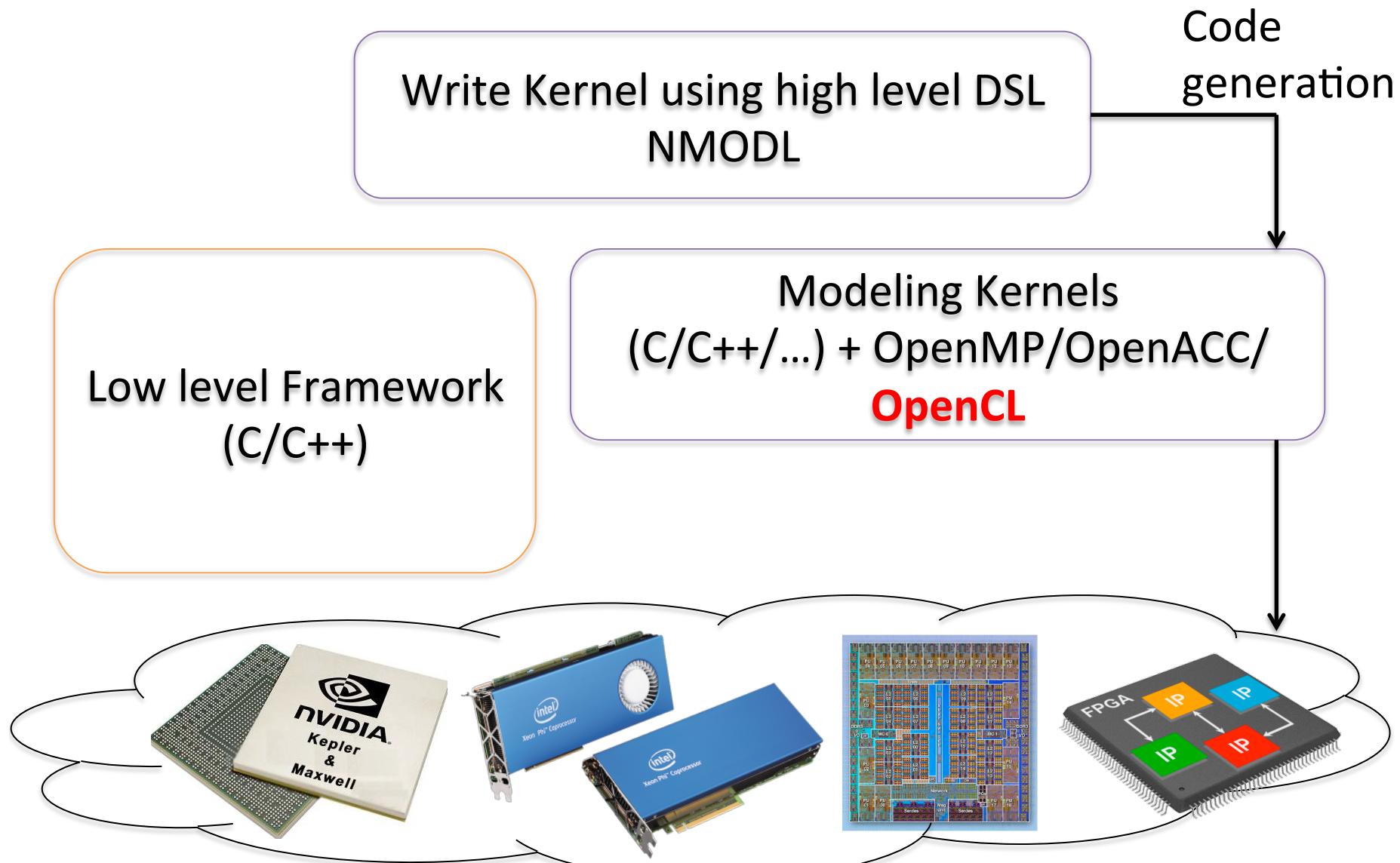
200k lines of code
All functionalities

**SaaS for Automated Performance Analysis/Modeling
of Annotated Kernels along with Performance History**

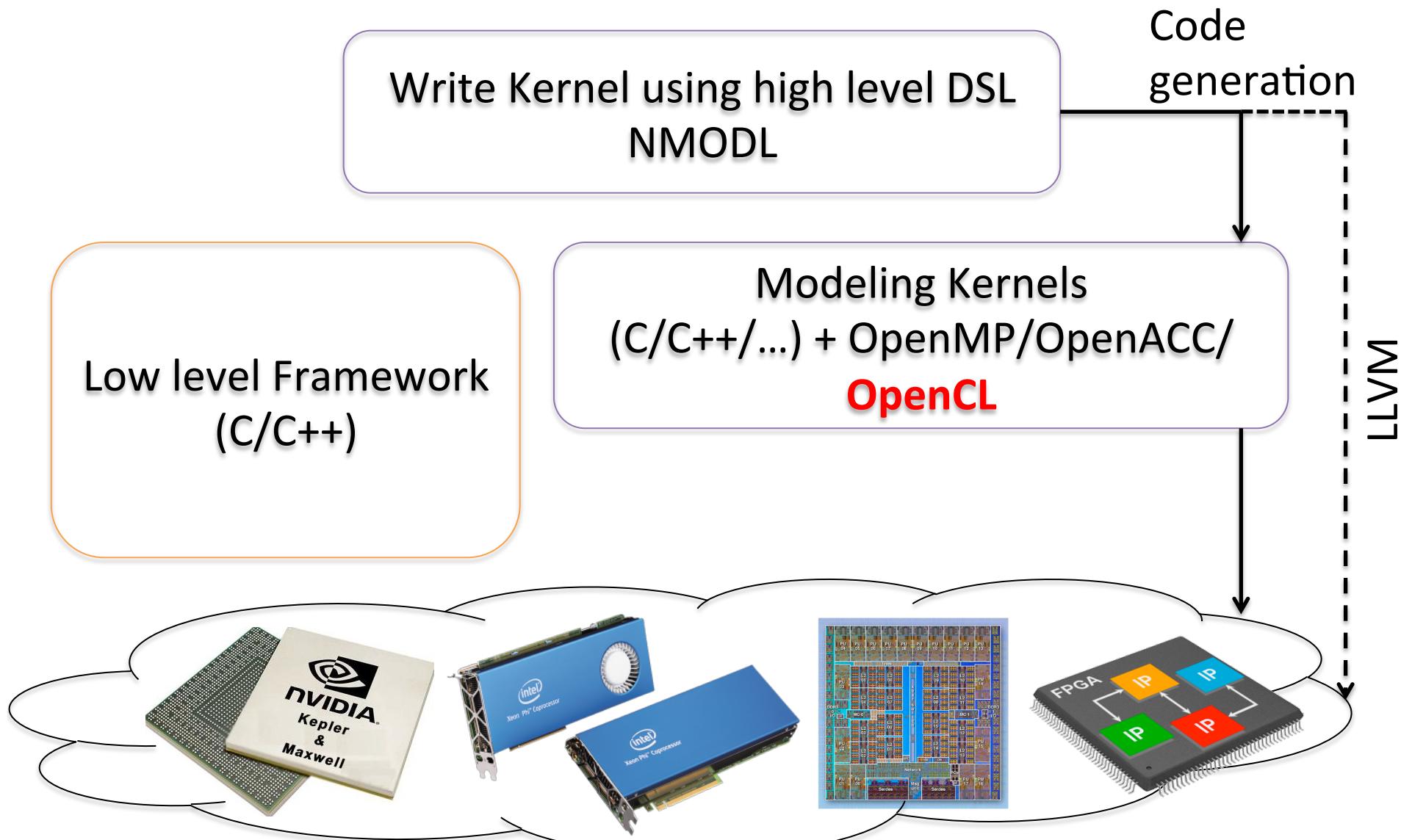
R&D - Development Workflow & Tools



R&D - Development Workflow & Tools

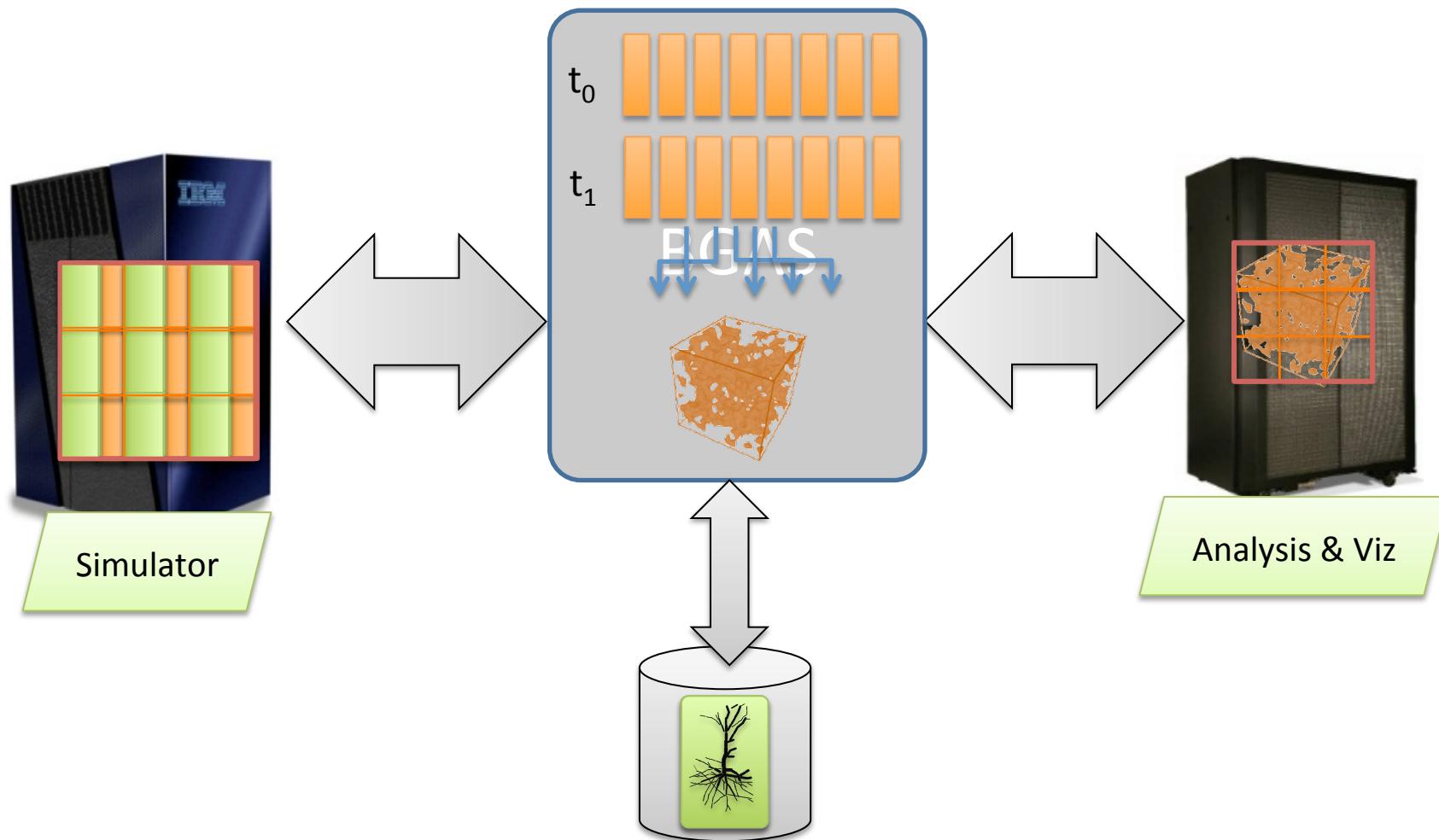


R&D - Development Workflow & Tools



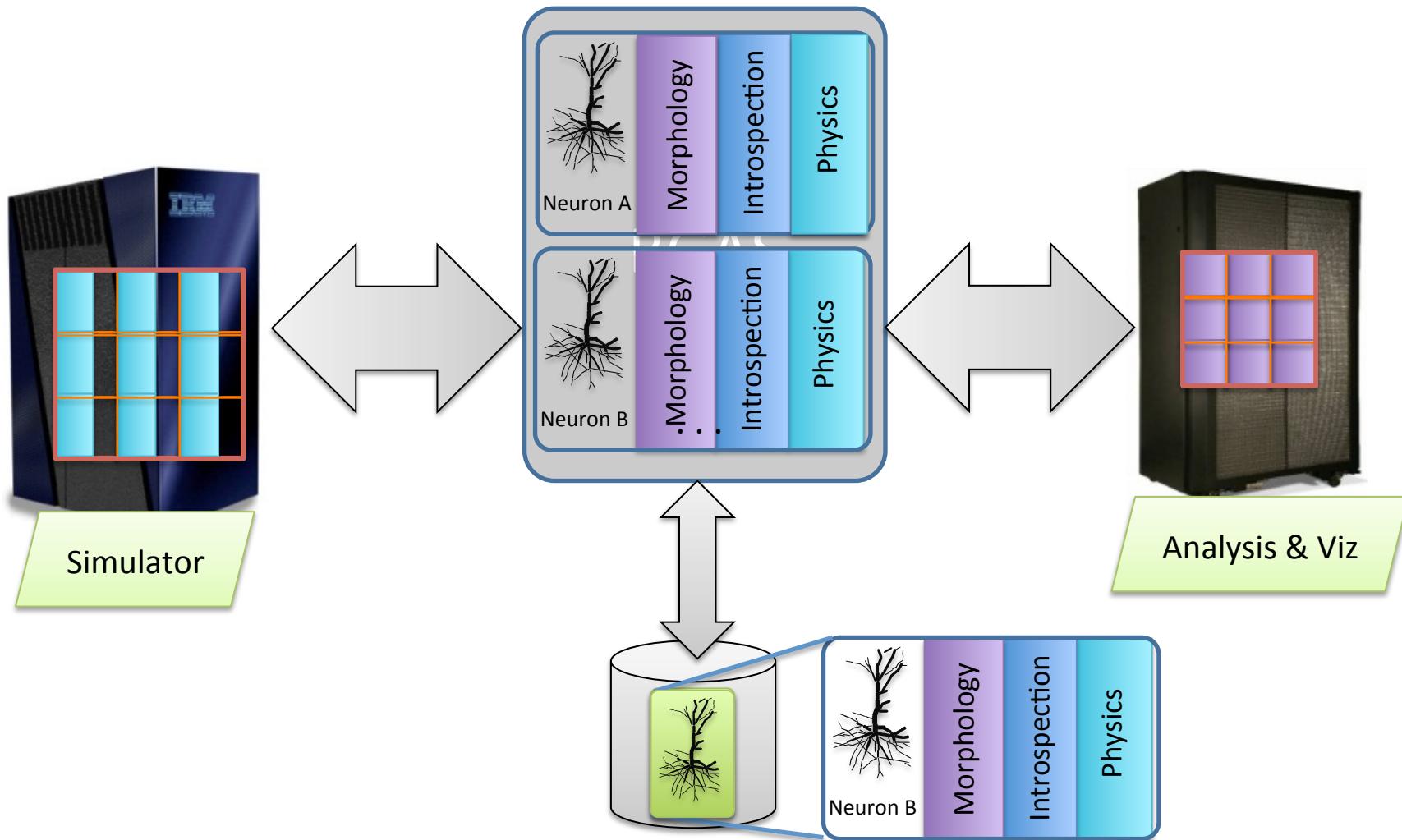
Deep Memory Hierarchy/Eco-System View

Blue Gene Active Storage (BGAS) – Active Result Buffer



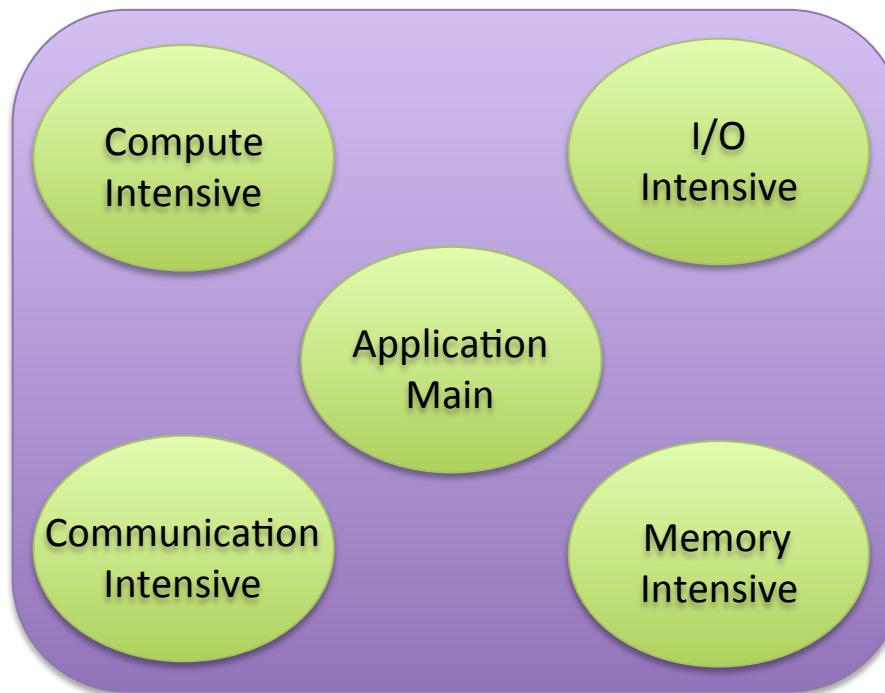
Deep Memory Hierarchy/Eco-System View

Blue Gene Active Storage (BGAS) – Model Container



Deep Memory Hierarchy/System View

Support Complex Workflows Minimizing Data Movement



Static Workflow

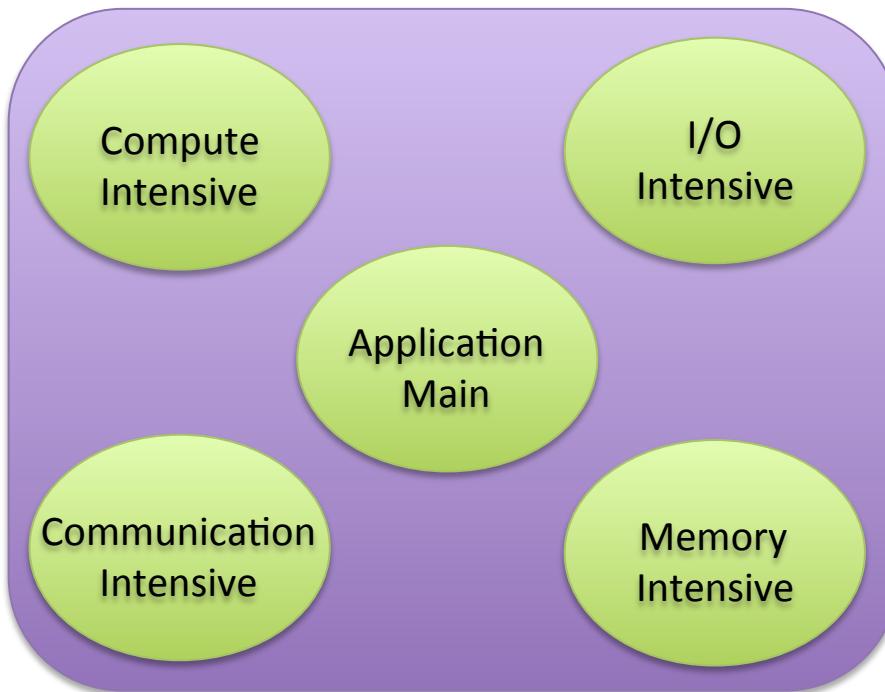
MPI/OpenMP for Building/Simulation

Spark for Analysis

All data stays on node

Deep Memory Hierarchy/System View

Support Complex Workflows Minimizing Data Movement



Static Workflow

MPI/OpenMP for Building/Simulation
Spark for Analysis
All data stays on node

Dynamic Workflow

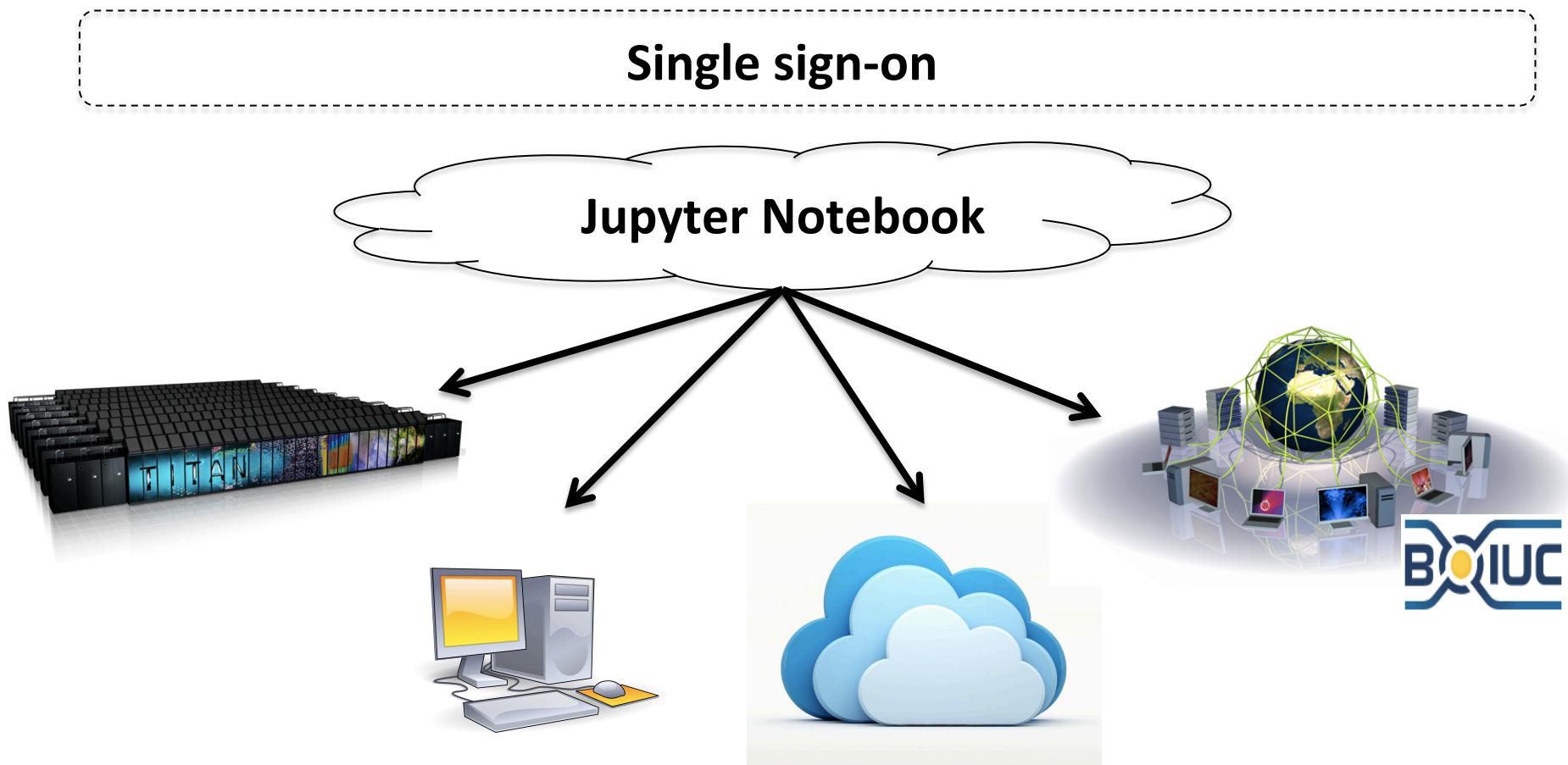
Simulation/Analysis/Visualization
competing for resources
On node run time / ZeroMq

Distributed Key Value Store

- Extend application memory by offering STL Map-like interface to additional stores
- Implementation should hide data coalescing, movement & prefetching to application
- Possible HPX run time integration to move data between various stores
- Wish to provide hints to compilers/run time

More information: Eilemann et al., Key-value Enabled Flash Memory Scientific Workflow with On-line Analysis and Visualization, 2016 IEEE International Parallel & Distributed Processing Symposium.

Increase Usability of Systems



Requires “Meta”-scheduler & Data Container+Nix

Hippocampus Preview

